

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:39 ; Search time 29.6 Seconds
(without alignments)
106.594 Million cell updates/sec

Title: US-08-930-480A-7
Perfect score: 56
Sequence: 1 KPSTPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriaph:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	470	11 Q7TMK1	Q7tmk1 mus musculus
2	46	82.1	356	11 Q924W1	Q924w1 rattus norv
3	46	82.1	738	11 Q92213	Q92213 rattus norv
4	46	82.1	1538	11 Q80U22	Q80u22 mus musculus
5	45	80.4	359	13 Q73756	Q73756 brachydanio
6	44	78.6	473	15 Q7W008	Q7w008 bordetella
7	44	78.6	766	12 Q9DW7	Q9dw7 rat cytomeg
8	43	76.8	199	3 Q875A9	Q875a9 podospora a
9	43	76.8	209	5 Q62173	Q62173 caenorhabdi
10	43	76.8	372	11 Q91WK4	Q91wk4 mus musculus
11	43	76.8	587	16 Q93H14	Q93h14 streptomyce
12	43	76.8	984	4 Q8HL12	Q8hl12 homo sapien
13	42	75.0	153	10 Q8LDP2	Q8ldp2 arabidopsis
14	42	75.0	153	10 Q9LPH1	Q9lph1 arabidopsis
15	42	75.0	189	10 Q84JX4	Q84jx4 oryza sativ
16	42	75.0	237	16 Q8FDM6	Q8fdm6 escherichia

17	42	75.0	254	16 Q8FT08	Q8ft08 corynebacte
18	42	75.0	330	10 Q9LQT4	Q9lqt4 arabidopsis
19	42	75.0	418	16 Q84876	Q84876 chlamydia t
20	42	75.0	592	5 Q9VZ49	Q9vz49 drosophila
21	42	75.0	676	12 Q7T9D9	Q7t9d9 sudan ebola
22	42	75.0	955	4 Q96DN2	Q96dn2 homo sapien
23	42	75.0	1201	5 Q868S2	Q868s2 anopheles g
24	42	75.0	1638	5 Q7YVM5	Q7yvm5 cryptospori
25	42	75.0	2838	5 Q8MP05	Q8mp05 tenebrio mo
26	41	73.2	96	17 Q972E4	Q972e4 sulfolobus
27	41	73.2	173	2 Q85951	Q85951 sphingomona
28	41	73.2	189	2 Q9L730	Q9l730 streptomyce
29	41	73.2	230	16 Q7U7W0	Q7u7w0 synecococc
30	41	73.2	330	4 Q96D28	Q96d28 homo sapien
31	41	73.2	339	5 Q9UL56	Q9ul56 leishmania
32	41	73.2	458	10 Q9FRJ1	Q9frj1 oryza sativ
33	41	73.2	458	10 Q7XCM0	Q7xcm0 oryza sativ
34	41	73.2	476	5 Q9VTU3	Q9vtu3 drosophila
35	41	73.2	582	16 Q82EA4	Q82ea4 streptomyce
36	41	73.2	737	13 Q90422	Q90422 brachydanio
37	41	73.2	750	13 Q9W633	Q9w633 cyprinus ca
38	41	73.2	762	13 Q9YIC6	Q9yic6 cyprinus ca
39	41	73.2	922	11 Q8BZX2	Q8bzx2 mus musculu
40	41	73.2	1984	12 Q9YQ45	Q9yq45 viral hemor
41	40	71.4	73	15 Q902F7	Q902f7 simian t-ly
42	40	71.4	73	15 Q8U17	Q8u17 simian t-ly
43	40	71.4	86	15 Q38316	Q38316 human immun
44	40	71.4	100	16 Q9FCM1	Q9fcm1 streptomyce
45	40	71.4	122	10 Q24523	Q24523 oryza sativ

ALIGNMENTS

RESULT 1
Q7TMK1
ID Q7TMK1 PRELIMINARY; PRT; 470 AA.
AC Q7TMK1;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZSCH II; TISSUE=Breast tumor;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K., Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh F., Hopkins R.F., Jordan H., Moore T., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettaman M., Madan A.C., Shchurko Y., Bouffard G.G., Whiting M., Madan A., Young A.C., Rodriguez S., Smit J., Gibbs R.A., Blakesley R.W., Touchman M., Madan A.C., Shevchenko Y., Bouffard G.G., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=CZSCH II; TISSUE=Breast tumor;

RA Strausberg R.;
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC055910; AAH55910.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 470 AA; 17277 MW; 6D90B4DF896BB090 CRC64;

Query Match 100.0%; Score 56; DB 11; Length 470;
 Best Local Similarity 100.0%; Pred. No. 0.57;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 242 PKPSTPPGSS 251

RESULT 2

Q924W1 ID Q924W1 PRELIMINARY; PRT; 356 AA.
 AC Q924W1; 2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE ALEX protein.
 GN ALEX
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95089824; PubMed=7997272;
 RA Kehlenbach R.H.; Huttner W.B.;
 RT "Xlas is a new type of G protein.";
 RL Nature 372:804-809(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Klemke M.; Kehlenbach R.H.; Huttner W.B.;
 RT "Two overlapping reading frames in a single exon encode interacting
 RT proteins - a novel way of gene usage."
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X84047; CAC39212.1; -;
 SQ SEQUENCE 356 AA; 37970 MW; 9849ABD0AE524A3D CRC64;

Query Match 82.1%; Score 46; DB 11; Length 356;
 Best Local Similarity 80.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 64 PKPSTPPGSS 73

RESULT 3

Q92213 ID Q92213 PRELIMINARY; PRT; 738 AA.
 AC Q92213; 1999 (TREMELrel. 10, Created)
 DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar;
 RA Wang Y.Z.; Kehlenbach R.H.; Huttner W.B.;
 RT "The XL-domain of rat Xlas is encoded by a single exon."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF093569; AAD03033.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 738 AA; 80340 MW; 51EA2B3A7D9D018A CRC64;

Query Match 82.1%; Score 46; DB 11; Length 738;
 Best Local Similarity 80.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 446 PKPSTPPGSS 455

RESULT 4

Q80U22 ID Q80U22 PRELIMINARY; PRT; 1538 AA.
 AC Q80U22; 2003 (TREMELrel. 24, Created)
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE MKIAA0375 protein (Fragment).
 GN MKIAA0375.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N.; Kikuno R.; Ohara R.; Inamoto S.; Aizawa H.; Yuasa S.;
 RA Nakajima D.; Nagase T.; Ohara O.; Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA genes:
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries."
 RL DNA Res. 10:35-48(2003).
 DR EMBL; AK122263; BAC65545.1; -;
 DR InterPro; IPR004012; Run.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF02759; RUN; 1.
 DR Pfam; PF00018; SH3; 1.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00593; RUN; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50826; RUN; 1.
 DR PROSITE; PS50002; SH3; 1.
 FT NON TER 1
 SQ SEQUENCE 1538 AA; 163862 MW; FBF93F32E3CD8EEE CRC64;

Query Match 82.1%; Score 46; DB 11; Length 1538;
 Best Local Similarity 80.0%; Pred. No. 73;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 1527 PTPSPPGSS 1536

RESULT 5

O73756 ID O73756 PRELIMINARY; PRT; 359 AA.
 AC O73756; 1998 (TREMELrel. 07, Created)
 DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
 DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Complement component Bfb (Fragment).
 GN BFB OR BFB.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KOC;

RX MEDLINE=99089896; PubMed=9874500;
 RA Gongora R., Figueroa F., Klein J.;
 RT "Independent duplications of Bf and C3 complement genes in the
 RL zebrafish."
 RL Scand. J. Immunol. 48:651-658(1998).
 DR EMBL: AF047412; AAC05096.1; -.
 DR ZFIN: ZDB-GENE-990415-34; bfb.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF00084; sushi; 2.
 DR SMART: PF00092; vwa; 1.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00327; VWA; 1.
 DR PROSITE: PS50234; VWF_A; 1.
 FT NON_TER 1 1
 FT NON_TER 359 359
 SQ SEQUENCE 359 AA; 40877 MW; 48770B63F310E10D CRC64;

Query Match 80.4%; Score 45; DB 13; Length 359;
 Best Local Similarity 80.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10

DB 60 PDPSVPPGSS 69

RESULT 6

Q7W008 PRELIMINARY; PRT; 473 AA.
 AC Q7W008;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN BP0030.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Chervach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica."
 RL Nat. Genet. 35:32-40(2003).
 RL EMBL: BX640411; CAE40409.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 473 AA; 50610 MW; EE84AED62A730A9F CRC64;

Query Match 78.6%; Score 44; DB 16; Length 473;
 Best Local Similarity 77.8%; Pred. No. 46;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9

DB 5 PPSSTPPGN 13

RESULT 7

Q9DWF7 PRELIMINARY; PRT; 766 AA.
 ID Q9DWF7

AC Q9DWF7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE PR34.
 GN R34.
 OS Rat cytomegalovirus (strain Maastricht).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Muromegalovirus.
 OX NCBI_TaxID=79700;
 RN [1]
 RP SEQUENCE OF 1-6 FROM N.A.
 RC STRAIN=Maastricht;
 RX MEDLINE=98139136; PubMed=9499096;
 RA Reisser P.S., Vink C., Van Dam J.G., Grauls G., Vanherle S.J.,
 RA Bruggeman C.A.;
 RT "The R33 G protein-coupled receptor gene of rat cytomegalovirus plays
 RT an essential role in the pathogenesis of viral infection."
 RL J. Virol. 72:2352-2363(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Maastricht;
 RX MEDLINE=20366325; PubMed=10906222;
 RA Vink C., Beuken E., Bruggeman C.A.;
 RT "Complete DNA sequence of the rat cytomegalovirus genome."
 RL J. Virol. 74:7656-7665(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Maastricht;
 RX MEDLINE=20473137; PubMed=11018281;
 RA Gruijthuijsen Y.K., Beuken E., Bruggeman C.A., Vink C.;
 RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
 RT spliced transcript."
 RL Virus Res. 69:119-130(2000).
 DR EMBL: AF232689; AAF99132.1; -.
 SQ SEQUENCE 766 AA; 84603 MW; 101ECS8097524704 CRC64;

Query Match 78.6%; Score 44; DB 12; Length 766;
 Best Local Similarity 70.0%; Pred. No. 75;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10

DB 154 PPSRPPGSA 163

RESULT 8

Q875A9 PRELIMINARY; PRT; 199 AA.
 ID Q875A9
 AC Q875A9;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to snRNP protein garl of Schizosaccharomyces pombe.
 OS Podospora anserina.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
 OX NCBI_TaxID=5145;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Genoscope;
 RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BX088700; CAD60706.1; -.
 DR InterPro: IPR007504; Garl.
 DR Pfam: PF04410; Garl; 1.
 SQ SEQUENCE 199 AA; 20050 MW; 7E7E85779A2B05F CRC64;

Query Match 76.8%; Score 43; DB 3; Length 199;
 Best Local Similarity 70.0%; Pred. No. 28;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10

DB 1 PPSSTPPGSS 10


```

AC Q8NH12;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Seven transmembrane helix receptor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tsubumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB065601; BAC05829.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004930; F:G-protein coupled receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007218; P:neuropeptide signaling pathway; IEA.
DR InterPro; IPR000832; GPCR secretin.
DR InterPro; IPR000203; PKD_Cys-rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF01825; GPs; 1.
DR PRINTS; PRO0249; GPCRSECRETIN.
DR SMART; SM00303; GPs; 1.
DR PROSITE; PS0221; GPs; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR Receptor; Transmembrane.
KW Receptor; Transmembrane.
SQ SEQUENCE 984 AA; 10440 MW; 01A789468A48B155 CRC64;

Query Match 76.8%; Score 43; DB 4; Length 984;
Best Local Similarity 70.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
DB 635 PSFGPPGSS 644

RESULT 13
Q8LDP2
ID Q8LDP2 PRELIMINARY; PRT; 153 AA.
AC Q8LDP2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY085879; AAM63092.1; -
KW Hypothetical protein.
SQ SEQUENCE 153 AA; 16475 MW; 730128E4C948C067 CRC64;

Query Match 75.0%; Score 42; DB 10; Length 153;
Best Local Similarity 77.8%; Pred. No. 31;

```

```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 9
DB 43 PQSPPPGSS 51

RESULT 14
Q9LPH1
ID Q9LPH1 PRELIMINARY; PRT; 153 AA.
AC Q9LPH1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE T3P20.13 protein (Hypothetical protein) (Atig53560).
GN T3P20.13 OR F22G10.10 OR Atig53560.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lee J.M., Vaysberg M., Sakano H., Lenz C., Liu S.X., Pham P.,
RA Toriumi M., Yu G., Chin C., Chio J., Choi E., Chung M., Gonzalez A.,
RA Howing B., Liu A., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC T3P20 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";

```

RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC018748; AAF78434.1; -
 DR EMBL; AC024260; AAG51969.1; -
 DR EMBL; AY099799; AAM20650.1; -
 DR EMBL; BT006546; AAP21354.1; -
 DR PIR; F96575; F96575.
 KW Hypothetical protein.
 SQ SEQUENCE 153 AA; 16479 MW; 321129E4D528DD0B CRC64;

Query Match 75.0%; Score 42; DB 10; Length 153;
 Best Local Similarity 77.8%; Pred. No. 31;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
 | : | | | | |
 Db 43 PQSPPPGS 51

RESULT 15

Q84JX4
 ID Q84JX4 PRELIMINARY; PRT; 189 AA.
 AC Q84JX4;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE OJ1372_D12.32 protein (OJ1372_D12.41 protein) (OJ1372_D12.50 protein).
 DE OJ1372_D12.32 OR OJ1372_D12.41 OR OJ1372_D12.50.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC clone:OJ1372_D12.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003827; BAC57668.1; -
 DR EMBL; AP003827; BAC57672.1; -
 DR EMBL; AP003827; BAC57676.1; -
 SQ SEQUENCE 189 AA; 20940 MW; 544F167C4AC7C49B CRC64;

Query Match 75.0%; Score 42; DB 10; Length 189;
 Best Local Similarity 70.0%; Pred. No. 38;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 10
 | : | | | | |
 Db 5 FWTTPPGA 14

Search completed: April 20, 2004, 10:27:03
 Job time : 30.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:04 ; Search time 8.4 Seconds
(without alignments)
61.988 Million cell updates/sec

Title: US-08-930-480A-7

Perfect score: 56

Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	329	1 GC3 MOUSE	P22436 mus musculus
2	56	100.0	398	1 GC3M MOUSE	P03987 mus musculus
3	43	76.8	732	1 TAU MOUSE	P10637 mus musculus
4	43	76.8	751	1 TAU RAT	P19332 rattus norv
5	42	75.0	237	1 YGHS ECOLI	Q46843 escherichia
6	41	73.2	441	1 GUN2 THEFU	P26222 thermomonos
7	40	71.4	386	1 SYTS HUMAN	O00445 homo sapien
8	40	71.4	476	1 TRZA RHOCO	O52725 rhodococcus
9	40	71.4	534	1 APG ARATH	P40602 arabidopsis
10	40	71.4	579	1 COR2 BRARE	O93375 brachydanio
11	40	71.4	592	1 FZD9 MOUSE	Q9r216 mus musculus
12	40	71.4	601	1 3BP1 MOUSE	F55194 mus musculus
13	40	71.4	1036	1 ACK1 HUMAN	Q07912 homo sapien
14	40	71.4	1055	1 ACK1 MOUSE	O54967 mus musculus
15	40	71.4	1167	1 SOR1 MOUSE	Q9jlca mus musculus
16	39.5	70.5	364	1 IE68 PRVKA	P24827 pseudorabie
17	39	69.6	181	1 RRP3 HORVU	O48609 hordeum vul
18	39	69.6	307	1 CC36 CAEBEL	P34803 caenorhabdi
19	39	69.6	386	1 SYTS MOUSE	Q9x0n5 mus musculus
20	39	69.6	386	1 SYTS RAT	P47861 rattus norv
21	39	69.6	489	1 OCLN POTTR	Q28793 potorous tr
22	39	69.6	532	1 GRB7 HUMAN	Q14451 homo sapien
23	39	69.6	536	1 GAG FSVMD	P03340 feline sarc
24	39	69.6	778	1 GELS CHICK	O93510 gallus gall
25	39	69.6	790	1 KIF9 MOUSE	Q9wy04 mus musculus
26	39	69.6	860	1 AREA PENRO	O13508 penicillium
27	39	69.6	865	1 NRPA PENUR	Q92269 penicillium
28	39	69.6	1167	1 SOR1 HUMAN	Q8wy21 homo sapien
29	39	69.6	3149	1 TEGU BEV	P03186 epstein-bar
30	38	67.9	126	1 DOCK HUMAN	O75956 homo sapien
31	38	67.9	240	1 PRA MYCTU	O53426 mycobacteri
32	38	67.9	349	1 ATF4 MOUSE	Q06507 mus musculus
33	38	67.9	353	1 ALC1_GORGO	P20758 gorilla gor

RESULT 1
GC3_MOUSE STANDARD; PRT; 329 AA.

ID GC3_MOUSE AC P22436;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 19 gamma-3 chain C region, secreted form.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]_TaxID=10090;

RP SEQUENCE FROM N.A.

RX MEDLINE=85027161; PubMed=6092053;

RA Wells J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,

RA Tucker P.W., Blattner F.R.;

RT "Structural analysis of the murine IgG3 constant region gene.";

RL EMBO J. 3:2041-2046(1984).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; J00451; ; NOT ANNOTATED_CDS.

CC PIR; B02156; G3MSC.

CC HSSP; P01957; 1FC1.

CC InterPro; IPR007110; Ig-like.

CC InterPro; IPR003597; Ig_cl.

CC InterPro; IPR003006; Ig_MHC.

CC Pfam; PF000047; Ig_3.

CC SMART; SM00407; IGc1; 2.

CC PROSITE; PS50835; IG_LIKE; 3.

CC PROSITE; PS00290; IG_MHC; 1.

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;

KW Transmembrane; Alternative splicing.

FT NON_TER 1

FT DOMAIN 1 97 CH1.

FT DOMAIN 98 113 HINGE.

FT DOMAIN 114 223 CH2.

FT DOMAIN 224 327 CH3.

SQ SEQUENCE 329 AA; 36228 MW; F45827174182BAD6 CRC64;

Query Match 100.0%; Score 56; DB 1; Length 329;

Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKPSTPPGSS 10

Db 101 PKPSTPPGSS 110

Tue Apr 20 13:50:56 2004

us-08-930-480a-7.rsp

```

RESULT 2
GC3M MOUSE STANDARD; PRT; 398 AA.
ID GC3M MOUSE
AC P0397;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG gamma-3 chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 6092053;
RX MEDLINE=85027161; PubMed=6092053;
RA Wells J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RT Tucker P.W., Blattner F.R.;
RA "Structural analysis of the murine IgG3 constant region gene.";
RL EMBO J. 3:2041-2046(1984).
RN [2]
RP SEQUENCE OF 328-398 FROM N.A.
RX MEDLINE=84041483; PubMed=6314258;
RA Komaromy M., Clayton L., Rogers J., Robertson S., Kettman J.,
RA Wall R.;
RT "The structure of the mouse immunoglobulin in gamma 3 membrane gene
segment.";
RL Nucleic Acids Res. 11:6775-6785(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00451; AAB59655.1; --
DR EMBL; V01526; CAA24767.1; ALT_SEQ.
DR PIR; A02156; G3MSM.
DR HSP; P01857; IFC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGCL; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane; Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 223 CH2.
FT DOMAIN 224 327 CH3.
FT TRANSMEM 346 362 POTENTIAL.
FT DOMAIN 363 398 CYTOPLASMIC (POTENTIAL).
FT CONFLICT 333 333 E -> G (IN REF. 2).
FT CONFLICT 342 342 E -> Q (IN REF. 2).
FT CONFLICT 388 388 P -> F (IN REF. 2).
SQ SEQUENCE 398 AA; 43929 MW; CF7F264B50A41B95 CRC64;
Query Match 100.0%; Score 56; DB 1; Length 398;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPSTPPGSS 10
DB 101 PKPSTPPGSS 110
RESULT 3
TAU MOUSE STANDARD; PRT; 732 AA.
ID TAU_MOUSE

```


CC Name=Tau-B;
 CC IsoId=PI0637-3; Sequence-VSP_003185, VSP_003187, VSP_003188,
 CC VSP_003189, VSP_003190;
 CC Name=Tau-C;
 CC IsoId=PI0637-4; Sequence-VSP_003185, VSP_003187, VSP_003188,
 CC VSP_003189;
 CC Name=Tau-D;
 CC IsoId=PI0637-5; Sequence-VSP_003185, VSP_003187, VSP_003188;
 CC Name=Tau-E;
 CC IsoId=PI0637-6; Sequence-VSP_003185, VSP_003186, VSP_003187,
 CC VSP_003188;
 CC
 CC !- TISSUE SPECIFICITY: EXPRESSED IN NEURONS AND AT A LOWER LEVEL IN
 CC THE LIVER AND KIDNEY. PNS-TAU IS EXPRESSED IN THE PERIPHERAL
 CC NERVOUS SYSTEM WHILE THE OTHERS ARE EXPRESSED IN THE CENTRAL
 CC NERVOUS SYSTEM.
 CC !- DEVELOPMENTAL STAGE: SHORTER FORMS OR LOW MOLECULAR WEIGHT TAU
 CC (LMW-TAU) ARE GENERALLY EXPRESSED AT EARLY DEVELOPMENT STAGES AND
 CC LONGER FORMS OR HIGH MOLECULAR WEIGHT TAU (HMW-TAU) IN THE ADULT
 CC BRAIN.
 CC !- DOMAIN: The tau/MAP repeat binds to tubulin. Type I isoforms
 CC contain 3 repeats while type II isoforms contain 4 repeats.
 CC !- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN
 CC S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK: CDC2,
 CC CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN
 CC MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY
 CC MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).
 CC !- DISEASE: May be involved in the pathogenesis of cytoplasmic
 CC inclusions (as Mallory bodies) in livers of mice chronically
 CC intoxicated with Griseofulvin or DDC (3,5-dithiohexanoyl-2,4-
 CC dihydrocollidine), a model for human alcoholic hepatitis.
 CC Alteration of tau (abnormal phosphorylation and crosslinking)
 CC could contribute to Mallory bodies formation and disturbance of
 CC microtubule function in alcoholic liver disease.
 CC !- SIMILARITY: Contains 4 Tau/MAP repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U12914; AA58343.1; -;
 CC EMBL; U12915; AA58344.1; -;
 CC EMBL; U12916; AA58345.1; -;
 CC EMBL; Z12133; CAA78121.1; -;
 CC EMBL; M93266; -; NOT ANNOTATED_CDS.
 CC EMBL; M18775; AAA40165.1; -;
 CC EMBL; M18776; AAA40166.1; -;
 CC EMBL; D30627; BAA18878.1; -;
 CC PIR; A28820; A28820.
 CC PIR; A45303; A45301.
 CC PIR; B28820; B28820.
 CC MGD; MGI:97180; Mapt.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC InterPro; IPR002955; Tau protein.
 CC InterPro; IPR001084; Tubulin Tau.
 CC Pfam; PF00418; tubulin-binding; 4.
 CC PRINTS; PR01261; TAUPROTEIN.
 CC PROSITE; PS00229; TAU MAP; 4.
 CC Microtubule; Cytoskeleton; Repeat; Alternative splicing; Acetylation;
 CC Phosphorylation.
 CC INIT MET 0 0 BY SIMILARITY.
 CC REPEAT 535 565 TAU/MAP MOTIF 1.
 CC REPEAT 566 596 TAU/MAP MOTIF 2.
 CC REPEAT 597 627 TAU/MAP MOTIF 3.
 CC REPEAT 628 659 TAU/MAP MOTIF 4.
 CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 CC DISULFID 582 613 BY SIMILARITY.
 CC VARSPLIC 33 90 Missing (in isoform Tau-B, isoform Tau-C,
 CC isoform Tau-D and isoform Tau-E).
 CC /FTid=VSP_003185.

FT VARSPLIC 91 112 Missing (in isoform Tau-E).
 FT /FTid=VSP_003186.
 FT VARSPLIC 113 349 Missing (in isoform Tau-A, isoform Tau-B,
 FT isoform Tau-C, isoform Tau-D and isoform
 FT Tau-E).
 FT /FTid=VSP_003187.
 FT Missing (in isoform Tau-A, isoform Tau-B,
 FT isoform Tau-C, isoform Tau-D and isoform
 FT Tau-E).
 FT /FTid=VSP_003188.
 FT Missing (in isoform Tau-B and isoform
 FT Tau-C).
 FT /FTid=VSP_003189.
 FT L -> KAALLSSQWNYSYHDLATITDLGL (in isoform
 FT Tau-B).
 FT /FTid=VSP_003190.
 FT D -> N (IN REF. 1).
 FT D -> N (IN REF. 1).
 FT P -> T (IN REF. 2; CAA78121).
 FT FT CONFLICT 527 527
 FT FT CONFLICT 671 671 E -> Q (IN REF. 1).
 SQ SEQUENCE 732 AA; 76112 MW; BFDFO767E41C7A3A CRC64;
 Query Match 76.8%; Score 43; DB 1; Length 732;
 Best Local Similarity 77.8%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPSTPPGS 9
 Db 468 PSFKTPGS 476
 RESULT 4
 ID TAU RAT STANDARD; PRT; 751 AA.
 AC P19332; Q63567; Q63677; Q9QW06;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 16-OCT-2003 (Rel. 40, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Microtubule-associated protein tau (Neurofibrillary tangle protein)
 DE (Paired helical filament-tau) (PHF-tau).
 DE MAPT OR MTAPT OR TAU
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM TAU-B).
 RX TISSUE=Phoeochromocytoma;
 RC MEDLINE=92179305; PubMed=1542696;
 RA Goedert M., Spillantini M.G., Crowther R.A.;
 RT "Cloning of a big tau microtubule-associated protein characteristic of
 RT the peripheral nervous system.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1983-1987(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM TAU-B).
 RC TISSUE=Dorsal root ganglion;
 RX MEDLINE=94013081; PubMed=8408300;
 RA Georgieff I.S., Liem R.K.H., Couchie D., Mavilia C., Nunez J.,
 RA Shelanski M.L.;
 RT "Expression of high molecular weight tau in the central and peripheral
 RT nervous systems.";
 RL J. Cell Sci. 105:729-737(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM TAU-F).
 RC STRAIN=Wistar; TISSUE=Brain;
 RX MEDLINE=94334997; PubMed=8057376;
 RA Sadot E., Marx R., Barg J., Behar L., Ginzburg I.;
 RT "Complete sequence of 3'-untranslated region of tau from rat central
 RT nervous system. Implications for mRNA heterogeneity.";
 RL J. Mol. Biol. 241:325-331(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS TAU-E AND TAU-G).
 RC TISSUE=Brain;


```

ID YGHS ECOLI STANDARD; PRT; 237 AA.
AC Q46843;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ATP-binding protein yghs.
GN YGHS OR B2985.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SIMILARITY: TO E.COLI YGHR AND YGHT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28377; AAA69152.1; -.
CC EMBL; AE000381; AAC76021.1; -.
CC PIR; G65084; G65084.
CC Ecogene; Egi3003; yghs.
KW Hypothetical protein; ATP-binding; Complete proteome.
FT NP_BIND 21 28 ATP (POTENTIAL).
SQ SEQUENCE 237 AA; 26346 MW; 69D8AE673D7DA6F CRC64;

Query Match 75.0%; Score 42; DB 1; Length 237;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPSTPPGS 9
Db 84 KPSTPPGN 91
|||||
|||||

RESULT 6
ID_GUN2_THEFU STANDARD; PRT; 441 AA.
AC P2622;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase E-2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-2)
DE (Cellulase E-2) (Cellulase E2).
GN CELB.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghargas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca.";
RL J. Bacteriol. 173:3397-3407(1991).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=YX;

```

```

RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RN Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 32-47.
RA Wilson D.B.;
RT "Cellulases of Thermomonospora fusca.";
RL Meth. Enzymol. 160:314-323(1988).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 32-317.
RX MEDLINE=94002001; PubMed=8399160;
RA Spezio M., Wilson D.B., Karplus P.A.;
RT "Crystal structure of the catalytic domain of a thermophilic
RT endocellulase.";
RL Biochemistry 32:9906-9916(1993).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC domain.
CC -!- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M73321; AAC06388.1; -.
CC PIR; A42360; A42360.
CC PIR; T12011; T12011.
CC PDB: 1TML; 31-JAN-94.
CC InterPro; IPR001919; Bac_celose-bind.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR001524; Glyco_hydro_6.
CC Pfam; PF00553; CBM_2; 1.
CC Pfam; PF01341; Glyco_hydro_6; 1.
CC PRINTS; PR00733; GLYHYDRLAS56.
CC ProDom; PD003733; Glyco_hydro_6; 1.
CC SMART; SM00637; CBD_II; 1.
CC PROSITE; PS00561; CBD_BACTERIAL; 1.
CC PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
CC PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 31
FT CHAIN 32 441 ENDOGLUCANASE E-2.
FT DOMAIN 32 320 CATALYTIC.
FT DOMAIN 321 340 LINKER.
FT DOMAIN 341 441 CELLULOSE-BINDING.
FT ACT_SITE 110 110
FT ACT_SITE 148 148
FT ACT_SITE 296 296
FT DISULFID 111 156
FT DISULFID 263 298
FT DISULFID 346 438
FT STRAND 36 36
FT TURN 40 41
FT TURN 43 50
FT HELIX 43 53
FT TURN 52 53
FT TURN 55 56
FT HELIX 57 63
FT TURN 64 66
FT STRAND 69 69
FT STRAND 71 73
FT HELIX 78 95
FT TURN 96 96
FT TURN 98 98
FT STRAND 100 103
FT STRAND 109 112
FT TURN 122 134
FT HELIX 135 135
FT TURN 135 135

```

```

FT STRAND 142 145
FT TURN 147 148
FT TURN 149 153
FT TURN 154 155
FT TURN 158 178
FT TURN 180 181
FT STRAND 183 187
FT TURN 196 205
FT TURN 206 207
FT TURN 208 211
FT STRAND 214 217
FT TURN 219 220
FT TURN 225 239
FT TURN 240 240
FT TURN 242 243
FT STRAND 245 249
FT TURN 254 255
FT TURN 259 260
FT TURN 266 267
FT STRAND 275 275
FT TURN 281 282
FT STRAND 283 288
FT TURN 292 293
FT STRAND 294 294
FT TURN 302 303
FT STRAND 305 305
FT TURN 307 315
FT TURN 316 316
SQ SEQUENCE 441 AA; 45843 MW; 8721854537092AE5 CRC64;

Query Match 73.2%; Score 41; DB 1; Length 441;
Best Local Similarity 70.0%; Pred. No. 72;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
Db 334 PPTPTPPGSS 343

RESULT 7
SYTS_HUMAN
ID SYTS_HUMAN STANDARD; PRT; 386 AA.
AC Q00445; Q86X72;
DC 01-NOV-1997 (Rel. 35, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DI 10-OCT-2003 (Rel. 42, Last annotation update)
DE Synaptotagmin V (SyTV).
GN SYTS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97321058; PubMed=9177789;
RA Craxton M.A., Olsen A., Goedert M.;
RT "Human synaptotagmin V (SYTS): sequence, genomic structure, and
chromosomal location.";
RL Genomics 42:165-169(1997).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-!- FUNCTION: May be involved in Ca(2+)-dependent exocytosis of
secretory vesicles through Ca(2+) and phospholipid binding to the
C2 domain or may serve as Ca(2+) sensors in the process of
vesicular trafficking and exocytosis (By similarity).
-!- SUBUNIT: Homodimer. Can also form heterodimers.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Synaptic
vesicles.
-!- SIMILARITY: Belongs to the synaptotagmin family.
-!- SIMILARITY: Contains 2 C2 domains.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X96783; CAA65579.1; -.
EMBL; BC046157; AAH46157.1; -.
HSSP; P21707; 1BYN.
Genew; HGNC:11513; SYTS.
DR MIM; 600782; -.
DR GO; GO:0007288; P:synaptic transmission; TAS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; LRI.
DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMIN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2 DOMAIN 1; 2.
DR PROSITE; PS00004; C2 DOMAIN 2; 2.
KW Transmembrane; Repeat; Synapse.
FT DOMAIN 1 24 VESICULAR (POTENTIAL).
FT TRANSMEM 25 45 POTENTIAL.
FT DOMAIN 46 386 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 124 211 C2 DOMAIN 1.
FT DOMAIN 253 344 C2 DOMAIN 2.
FT CONFLICT 111 111 R -> Q (IN REF. 2).
SQ SEQUENCE 386 AA; 42900 MW; 96A36792D177FD55 CRC64;

Query Match 71.4%; Score 40; DB 1; Length 386;
Best Local Similarity 70.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
Db 11 PSPDTPPDSS 20

RESULT 8
TRZA RHOCO
ID TRZA RHOCO STANDARD; PRT; 476 AA.
AC Q52725;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-triazine hydrolase (EC 3.8.1.-) (N-ethylammelane chlorohydrolase).
DR TRZA.
GN Rhodococcus corallinus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```



```

QY 1 PKFSTPPGSS 10
Db 67 PKFVAPPGPS 76

RESULT 10
COE2-BRARE STANDARD; PRT; 579 AA.
AC O93375;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription factor COE2.
GN COE2.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99002949; PubMed=9784615;
RA Bally-Cuif L., Dubois L., Vincent A.;
RT "Molecular cloning of Zco2, the zebrafish homolog of Xenopus Xco2
and mouse EBF-2, and its expression during primary neurogenesis.";
RL Mech. Dev. 77:85-90(1998).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC it covers the anterior neural plate. Widely expressed in the
CC presumptive mesencephalon and rhombomeres 1-4 until the 2-3-somite
CC stage, with expression persisting in ng1-5-positive clusters. First
CC detected in the olfactory placodes at the 5-somite stage. In the
CC spinal cord, detected in ng1-positive clusters of primary
CC neuroblasts during the early somite stages. Expression decreases
CC in the spinal cord from the 30-somite stage but persists in the
CC olfactory bulb and regions of the rhombencephalon and brain.
CC -1- SIMILARITY: Belongs to the COE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF072657; AAC96103.1; -.
DR ZFIN; ZDB-GENE-990715-11; coe2.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 1.
DR SMART; SM00429; IPT; 1.
DR SMART; SMO0353; HLH; 1.
DR PROSITE; PS01345; COE; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Developmental protein; Zinc-finger.
FT ZN_FING 149 168 C5-TYPE (POTENTIAL).
FT DOMAIN 460 537 SER/THR/PRO-RICH.
SQ SEQUENCE 579 AA; 63528 MW; E2DD1919055EEA43 CRC64;

Query Match 71.4%; Score 40; DB 1; Length 579;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PSTPPGSS 10
Db 517 PSSPPGSS 524

RESULT 11
FZD9_MOUSE
```

CC craniofacial regions and nephric ducts. In the adult, expression
CC is abundant in heart, brain, testis and skeletal muscle. In the
CC testis, expressed in all spermatogenic cell types. Lower levels in
CC adult lung, liver and kidney. Barely detectable in spleen.
CC Expressed also in chondrocytes.
CC -!- DEVELOPMENTAL STAGE: Not detected at embryonic day 7 (E7), weakly
CC at E11 and strongly at E15 and E17. Expression covers the entire
CC neural tube at 9.5 dpc, decreases at 10.5 dpc and becomes
CC detectable only in the lumbar to tail regions at 11.5 dpc. In the
CC somites, expression begins at 10.5 dpc to become upregulated all
CC along the rostrocaudal trunk axis at 11.5 dpc. In craniofacial
CC territories, expression is first detected at 11.5 dpc in restricted
CC areas of the nose, the maxillar mandibular and second branchial
CC arch anlagen. At 11.5 dpc, predominantly expressed in restricted
CC areas of the nose, dorsally to the eye and in the caudal
CC pharyngeal region.
CC -!- DOMAIN: Lys-Thr-X-X-Trip motif is involved in the activation of
CC the Wnt/beta-catenin signaling pathway (By similarity).
CC -!- DOMAIN: The FZ domain is involved in binding with Wnt ligands (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor
CC family.
CC -!- CAUTION: Has been first described as FZD3 in litterature.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF088850; RAD27789.1; -
CC EMBL; AK021164; BAB32311.1; -
CC EMBL; AF033585; AAB87503.2; -
CC EMBL; X17709; CAB44237.1; -
CC MGD; MGI:1313278; Fzd9.
CC InterPro; IPR000539; Frizzled.
CC InterPro; IPR000024; Fz domain.
CC Pfam; PF01534; Frizzled; GPCR secretin.
CC Pfam; PF01392; Fz; 1
CC PRINTS; PR00489; FRIZZLED.
CC SMART; SM00063; FRI; 1.
CC PROSITE; PS50038; FZ; 1.
CC PROSITE; PS50261; G-PROTEIN RECEPTOR; Transmembrane;
CC Multigene family; G-protein coupled receptor; Glycoprotein; Signal.
CC Developmental protein; Wnt signaling pathway; Signal.
CC SIGNAL 1 23
CC CHAIN 24 592
CC DOMAIN 24 230
CC TRANSMEM 231 251
CC DOMAIN 252 267
CC TRANSMEM 268 288
CC DOMAIN 289 316
CC TRANSMEM 317 337
CC DOMAIN 338 356
CC TRANSMEM 357 377
CC DOMAIN 378 401
CC TRANSMEM 402 422
CC DOMAIN 423 448
CC TRANSMEM 449 469
CC DOMAIN 470 509
CC TRANSMEM 510 530
CC DOMAIN 531 592
CC DOMAIN 593 666
CC SITE 533 538
CC CARBOHYD 54 54
CC CARBOHYD 159 159
CC CONFLICT 66 66
CC CONFLICT 73 74
CC CONFLICT 93 93
CC LYS-THR-X-X-TRIP MOTIF
CC N-LINKED (GLNAC...) (POTENTIAL).
CC N-LINKED (GLNAC...) (POTENTIAL).
CC S -> P (IN REF. 3).
CC QL -> HC (IN REF. 2).
CC L -> F (IN REF. 2).

FT CONFLICT 144 144 P -> S (IN REF. 4).
FT CONFLICT 221 221 E -> K (IN REF. 4).
FT CONFLICT 237 237 A -> P (IN REF. 4).
FT CONFLICT 308 308 G -> D (IN REF. 3).
FT CONFLICT 374 374 V -> F (IN REF. 4).
FT CONFLICT 592 592 L -> P (IN REF. 2).
SQ SEQUENCE 592 AA; 64994 MW; 21B2D4F8CE232965 CRC64;

Query Match 71.4%; Score 40; DB 1; Length 592;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 PKPSTPPGSS 10
|:|:|:|:
Db 178 PRPARPPGDS 187

RESULT 12
ID 3BP1 MOUSE STANDARD; PRT; 601 AA.
AC P55194; Q59KK8;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE SH3-domain binding protein 1 (3BP-1).
GN SH3BP1 OR 3BP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95347339; PubMed=7621827;
RA Cicchetti P., Ridley A.J., Zheng Y., Cerione R.A., Baltimore D.;
RT "3BP-1, an SH3 domain binding protein, has GAP activity for Rac and
RT inhibits growth factor-induced membrane ruffling in fibroblasts.";
RL EMBO J. 14:3127-3135(1995).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=23388257; PubMed=12477932;
RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler K.G.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]
RP SEQUENCE OF 263-601 FROM N.A.
RX MEDLINE=92358242; PubMed=1379745;
RA Cicchetti P., Mayer B.J., Thiel G., Baltimore D.;
RT "Identification of a protein that binds to the SH3 region of Abl and
RT is similar to Bcr and GAP-rho.";
RL Science 257:803-806(1992).
CC -!- FUNCTION: Binds differentially to the SH3 domains of certain
CC proteins of signal transduction pathways. This protein binds
CC preferentially to c-Abl proto-oncogene, SRC and GRB2. Shows GAP
CC activity for Rac-related proteins but not for Rho- or Ras-related
CC proteins. It inhibits PDGF-induced membrane ruffling mediated by
CC Rac.

Raba S.S., Loquellano N.A., Petersen G.J., Abramson R.D., Mullany S.J.,
Posak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywiński M.I., Skalska U., Smalhus D.E.,
Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[3]
STRUCTURE BY NMR OF 448-489.
MEDLINE-99287324; PubMed-10160579;
Mott H.R., Owen D., Nietlispach D., Lowe P.N., Manser E., Lim L.,
Laue E.D.;
"Structure of the small G protein Cdc42 bound to the GTPase-binding
domain of ACK.";
Nature 399:384-388 (1999).
CC -1- FUNCTION: Tyrosine kinase, that after binding to CDC42, inhibits
both its intrinsic and stimulated GTPase activity.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphates.
CC -1- SUBUNIT: Interacts with CDC42.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId-Q07912-1; Sequence=Displayed;
CC Name=2;
CC IsoId-Q07912-2; Sequence=VSP_008655, VSP_008656;
CC Note=No experimental confirmation available;
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -1- SIMILARITY: Contains 1 CRIB domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L13738; AAA53570.2; - -
DR EMBL; BC008884; AAH08884.1; - -
DR EMBL; BC028164; AAH28164.1; - -
DR PIR; S33596; S33596.
DR PDB; 1CF4; 27-JUN-01.
DR MIM; 606994; - -
DR GO; GO:0005095; P:GTPase inhibitor activity; TAS.
DR GO; GO:0004715; P:non-membrane spanning protein tyrosine kinase; - ; TAS.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; TAS.
DR InterPro; IPR0000719; Prot_kinase.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00627; UBA; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50108; CRIB; FALSE NEG
KW Transferase; Tyrosine-protein kinase; ATP-binding; SH3 domain;
KW 3D-structure; Alternative splicing.

FT DOMAIN 126 385 PROTEIN KINASE.
 FT DOMAIN 386 448 SH3.
 FT DOMAIN 454 466 CRIB.
 FT DOMAIN 577 956 PRO-RICH.
 FT NP BIND 132 140 ATP (BY SIMILARITY).
 FT BINDING 158 158 ATP (BY SIMILARITY).
 FT ACT SITE 252 252 BY SIMILARITY.
 FT VARSPLIC 485 528 LYLGNPDPPDLISVELSTRPPHGLGVKKPTVDPSVSDQ
 FT DPL -> CPFSAPFCHPAPETCGVLWTGRACADPRL
 FT HVSSTRKGL (in isoform 2).
 FT /FTID=VSP_008655.
 FT Missing (in isoform 2).
 FT /FTID=VSP_008656.
 FT G -> V (IN REF. 2; AAH08884).
 FT TRTSHASDTWMPGVTIWEMTYCQEPWGLNGSLQILHKID
 FT KEGRLR -> PWRDISASSSTQFHPVPCFPFTSLAKL
 FT LLRHVSVPASSREIKLVSLIC (IN REF. 2;
 FT AAH08884).
 FT Missing (IN REF. 2; AAH08884).
 FT CONFLICT 353 1036 MW; B9B90BA7E3E22DFF CRC64;
 FT SEQUENCE 1036 AA; 114327 MW; 71.4%; Score 40; DB 1; Length 1036;
 Query Match Best Local Similarity 70.0%; Pred. NO. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 PKPSTPPGSS 10
 Db 791 PSLVPPGSS 800

RESULT 14
 ACK1 MOUSE
 ID ACK1 MOUSE STANDARD; PRT; 1055 AA.
 AC O54967; Q8C2U0; Q8K0K4;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Activated CDC42 kinase 1 (EC 2.7.1.112) (ACK-1) (Non-receptor protein
 DE tyrosine kinase Ack) (Tyrosine kinase non-receptor protein 2).
 GN ACK1 OR TNK2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Her J.-H., Bolen J.B.;
 RT "The protein tyrosine kinase Ack is associated with and activated in
 RT vivo by CDC42Hs.";
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC STRAIN=C57BL/6; TISSUE=Brain, and Colon;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshitake S., Carninci P., Prange C.,
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RN SEQUENCE OF 121-1055 FROM N.A. (ISOFORM 3).
 RP STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schirni L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Jackson I.J., Jarvis E.D.,
 RA Grimmond S., Gustincich S., Hirokawa N., Gough J.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 CC -!- FUNCTION: Tyrosine kinase, that after binding to CDC42, inhibits
 CC both its intrinsic and stimulated GTPase activity (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: Interacts with CDC42 (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=O54967-1; Sequence=Displayed;
 CC Note=No experimental confirmation available;
 CC Name=2;
 CC IsoId=O54967-2; Sequence=VSP_008657, VSP_008658;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=O54967-3; Sequence=VSP_008657;
 CC -!- SIMILARITY: Belongs to the Tyr family of protein kinases.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -!- SIMILARITY: Contains 1 CRIB domain.
 CC -!- CAUTION: Ref.2 (AAH31168) sequence differs from that shown due to
 CC the presence of a sequence of unknown origin.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF037260; AAC04786.1; -;
 CC EMBL; BC031168; AAH31168.1; ALT_SEQ.
 CC EMBL; BC052421; AAH52421.1; -;
 CC EMBL; AK087965; BAC40063.1; -;
 CC HSSP; P11362; 1FGK.
 CC MGD; MGI:1858308; Tnk2.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR001452; SH3.
 CC InterPro; IPR001245; Tyr_kinase.

DR InterPro; IPR008266; Tyr_pkinase_AS.
DR InterPro; IPR000449; UBA domain.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00627; UBA; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_Kinase; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00002; SH3; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; SH3 domain;
FT DOMAIN 126 385 PROTEIN KINASE.
FT DOMAIN 386 448 SH3.
FT DOMAIN 454 466 CRIB.
FT DOMAIN 517 950 PRO-RICH.
FT NP_BIND 132 140 ATP (BY SIMILARITY).
FT BINDING 158 158 ATP (BY SIMILARITY).
FT ACT_SITE 252 232 BY SIMILARITY.
FT VARSPPLIC 515 531 RPPPPPPPPAIFTQKT -> KP (in isoform 2 and isoform 3).
FT /FTID=VSP_008657.
FT Missing (in isoform 2).
FT /FTID=VSP_008658.
FT SG -> RR (IN REF. 2).
FT V -> A (IN REF. 2 AND 3).
FT L -> V (IN REF. 3).
FT A -> T (IN REF. 2; AAHS2421 AND 3).
SQ SEQUENCE 1055 AA; 116835 MW; FBC9DAC85B2003F CRC64;
Query Match 71.4%; Score 40; DB 1; Length 1055;
Best Local Similarity 70.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 PKPSTPPGSS 10
Db 807 PSLVPPGSS 816
RESULT 15
SOR1_MOUSE STANDARD; PRT; 1167 AA.
AC Q9JLC4; Q8V745; Q92211; Q9QV21;
DT 28-FEB-2003 (Rel. 41, Created).
DT 28-FEB-2003 (Rel. 41, Last sequence update).
DT 10-OCT-2003 (Rel. 42, Last annotation update).
DE VPS10 domain-containing receptor SorCS1 precursor (mSorCS).
GN SORCS OR SORCS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=20086769; PubMed=10600506;
RA Hermey G., Riedel I.B., Hampe W., Schaller H.C., Hermans-Borgmeyer I.;
RT "Identification and characterization of SorCS, a third member of a novel receptor family.";
RL Biochem. Biophys. Res. Commun. 266:347-351(1999).
RN [2]
RC SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=20225481; PubMed=10760602;
RA Hermey G., Schaller H.C.;
RT "Alternative splicing of murine SorCS leads to two forms of the receptor that differ completely in their cytoplasmic tails.";
RL Biochim. Biophys. Acta 1491:350-354(2000).
RN [3]
RC SEQUENCE FROM N.A. (ISOFORM 3).

RC STRAIN=C57BL/6J;
RA Hermey G.;
RT "A third splice variant of mSorCS.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 424-1167 FROM N.A. (ISOFORM 4).
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo A.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=sorCSb;
CC IsoId=Q9JLC4-1; Sequence=Displayed;
CC Name=2; Synonyms=sorCSa;
CC IsoId=Q9JLC4-2; Sequence=VSP_006205;
CC Name=3; Synonyms=sorCSc;
CC IsoId=Q9JLC4-3; Sequence=VSP_006206;
CC Name=4;
CC IsoId=Q9JLC4-4; Sequence=VSP_006207;
CC -!- TISSUE SPECIFICITY: Isoform 1 is highly expressed in brain, and at lower levels in heart, liver and kidney. Detected in newborn brain and in adult olfactory bulb and cerebral cortex. Isoform 2 is highly expressed in liver, and at lower levels in heart, brain, kidney and testis.
CC -!- SIMILARITY: Contains 1 PKD domain.
CC -!- SIMILARITY: Contains 5 BNR repeats.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a frameshift in position 472.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF137367; AAF24748.1; -
CC EMBL; AF195056; AAF68196.1; -
CC EMBL; AF284755; AAL56666.1; -
CC EMBL; BC007486; AAH07486.1; ALT_FRAME.
CC MGD; MGI:1329666; Sorcs.
CC InterPro; IPR002860; GH_BNR.
CC InterPro; IPR00601; PKD.
CC InterPro; IPR006581; VPS10.
CC Pfam; PF02012; BNR; 5.
CC Pfam; PF00801; PKD; 1.
CC SMART; SM00089; PKD; 2.
CC SMART; SM00602; VPS10; 1.
CC PROSITE; PS00093; PKD; 1.
KW Signal; Transmembrane; Repeat; Alternative splicing.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1167 VPS10 DOMAIN-CONTAINING RECEPTOR SORCS1.
FT DOMAIN 34 1098 LUMENAL (POTENTIAL).

```
FT TRANSMEM 1099 1119 POTENTIAL.
FT DOMAIN 1120 1167 CYTOPLASMIC (POTENTIAL).
FT REPEAT 208 219 BNR 1.
FT REPEAT 256 267 BNR 2.
FT REPEAT 492 503 BNR 3.
FT REPEAT 569 580 BNR 4.
FT REPEAT 611 622 BNR 5.
FT DOMAIN 802 893 PKD.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 775 775 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 815 815 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 846 846 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 928 928 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1124 1167 RVALPSPSPSAQGDSSRLRQRPATPPSPKRGSGAQ
FT isoform 2).
FT /FTid=VSP 006205.
FT RVALPSPSPSAQGDSSRLRQRPATPPSPKRGSGAQ
FT FAI -> KIPGINVYAQMNEKEQLINPVSHSRSPVPH
FT PDLRRPGQLVDEKVESQLLKG (in isoform 3).
FT /FTid=VSP 006206.
FT RVALPSPSPSAQGDSSRLRQRPATPPSPKRGSGAQ
FT FAI -> KIPGINVYAQMNEKEQLINPVSHSRSPVPH
FT PDLRRPGQLVDEKVESQLLGE (in isoform 4).
FT /FTid=VSP 006207.
FT S -> Y (IN REF. 1).
FT Q -> QG (IN REF. 4).
FT A -> T (IN REF. 4).
SQ SEQUENCE 1167 AA; 129631 MW; 3563396487105B73 CRC64;
```

Query Match 71.4%; Score 40; DB 1; Length 1167;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
|:|:|:|
Db 1147 PRPATPPSS 1155

Search completed: April 20, 2004, 10:25:37
Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:22:14 ; Search time 10.8 seconds
(without alignments)
89.066 Million cell updates/sec

Title: US-08-930-480A-7

Perfect score: 56

Sequence: 1 PKSTPPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	329	1 G3MSC	Ig gamma-3 chain C
2	56	100.0	398	1 G3MSM	Ig gamma-3 chain C
3	43	76.8	209	2 T20975	hypothetical prote
4	43	76.8	341	2 B28820	microtubule-associ
5	43	76.8	364	2 A28820	microtubule-associ
6	43	76.8	374	2 S46264	microtubule-associ
7	43	76.8	432	2 J50306	microtubule-associ
8	43	76.8	686	2 A38235	microtubule-associ
9	43	76.8	733	2 A45301	microtubule-associ
10	43	76.8	1212	2 S27771	RNA-directed DNA p
11	42	75.0	153	2 F96575	hypothetical prote
12	42	75.0	237	2 G65084	hypothetical prote
13	42	75.0	330	2 D96787	protein T4012.3 [i
14	42	75.0	418	2 D71460	probable membrane
15	41	73.2	173	2 T31243	hypothetical prote
16	41	73.2	339	2 T46713	hypothetical prote
17	41	73.2	441	2 T12011	cellulase (EC 3.2.
18	40	71.4	122	2 T04366	probable peroxidase
19	40	71.4	190	2 T35570	hypothetical prote
20	40	71.4	229	2 T52364	hypothetical prote
21	40	71.4	235	2 A72594	hypothetical prote
22	40	71.4	274	2 T48819	hypothetical prote
23	40	71.4	463	2 T39004	probable histone h
24	40	71.4	477	2 T46666	N-ethylamine ch
25	40	71.4	534	2 S21961	proline-rich prote
26	40	71.4	601	2 S56144	SH3 domain binding
27	40	71.4	1014	2 A55260	cytotoxic necrotiz
28	40	71.4	1091	2 S33596	protein-tyrosine k
29	40	71.4	1137	2 A86335	T20H2.9 protein -

ALIGNMENTS

RESULT 1

G3MSC

Ig gamma-3 chain C region, secreted form - mouse

C;Species: Mus musculus (house mouse)

C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999

C;Accession: B02156

R;Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatt

EMBO J. 3, 2041-2046, 1984

A;Title: Structure analysis of the murine IgG3 constant region gene.

A;Reference number: A02156; MUID:85027161; PMID:6092053

A;Accession: B02156

A;Molecule type: DNA

A;Residues: 1-329 <WEL>

A;Cross-references: GB:J00451

A;Note: the sequence was determined from the germline gene

C;Genetics:

A;Introns: 97/1; 113/1; 223/1

C;Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C;Superfamily: immunoglobulin C region; immunoglobulin homology

C;Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul

F;19-83/Domain: immunoglobulin homology <IM1>

F;97-112/Region: hinge

F;136-205/Domain: immunoglobulin homology <IM2>

F;242-309/Domain: immunoglobulin homology <IM3>

F;179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 56; DB 1; Length 329;

Best Local Similarity 100.0%; Pred. No. 0.57;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKSTPPGSS 10

Db 101 PKSTPPGSS 110

RESULT 2

G3MSM

Ig gamma-3 chain C region, membrane-bound form - mouse

C;Species: Mus musculus (house mouse)

C;Date: 13-Aug-1986 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999

C;Accession: A02156; A02155

R;Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatt

EMBO J. 3, 2041-2046, 1984

A;Title: Structure analysis of the murine IgG3 constant region gene.

A;Reference number: A02156; MUID:85027161; PMID:6092053

A;Accession: A02156

A;Molecule type: DNA

A;Residues: 1-398 <WEL>

A;Cross-references: GB:J00451; NID:g194392; PIDN:AAB59655.1; PID:g194433

A;Note: the sequence was determined from the germline gene

R;Komaromy, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.

Nucleic Acids Res. 11, 6775-6785, 1983
 A:Title: The structure of the mouse immunoglobulin in gamma-3 membrane gene segment.
 A:Reference number: A02155; MUID:84041483; PMID:6314258
 A:Accession: A02155
 A:Molecule type: DNA
 A:Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>
 A:Cross-references: GB:K00688
 A>Note: the sequence was determined from the germline gene
 C:Genetics:
 A:Introns: 97/1; 113/1; 223/1; 328/1; 371/3
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger disulfide bonds.
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin
 F:19-83/Domain: immunoglobulin homology <IM1>
 F:97-112/Region: hinge
 F:136-205/Domain: immunoglobulin homology <IM2>
 F:242-309/Domain: immunoglobulin homology <IM3>
 F:346-362/Domain: transmembrane #status predicted <TM>
 F:363-398/Domain: intracellular #status predicted <INT>
 F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 56; DB 1; Length 398;
 Best Local Similarity 100.0%; Pred. No. 0.68;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 101 PKPSTPPGSS 110

RESULT 3
 220975
 hypothetical protein F15D3.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20975
 R:White, S.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19353
 A:Accession: T20975
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-209 <WIL>
 A:Cross-references: EMBL:Z81063; PIDN:CAB02955.1; GSPDB:GN00019; CESP:F15D3.6
 A:Experimental source: clone F15D3
 C:Genetics:
 A:Gene: CESP:F15D3.6
 A:Map position: 1
 A:Introns: 11/2; 66/3; 95/3; 153/3; 175/3

Query Match 76.8%; Score 43; DB 2; Length 209;
 Best Local Similarity 80.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 DB 199 PKPSTPPGSS 208

RESULT 4
 B28820
 microtubule-associated protein tau type 2 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
 C:Accession: B28820
 R:Lee, G.; Cowan, N.; Kirschner, M.
 Science 239, 285-288, 1988
 A:Title: The primary structure and heterogeneity of tau protein from mouse brain.
 A:Reference number: A94298; MUID:88099510; PMID:3122323
 A:Accession: B28820
 A:Molecule type: mRNA
 A:Residues: 1-341 <LEE>

A:Cross-references: GB:M18775; NID:g201114; PIDN:AAA40165.1; PID:g201115
 C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
 C:Keywords: alternative splicing; microtubule binding; tandem repeat
 F:183-213/Domain: MAP2/tau repeat homology <MT1>
 F:214-244/Domain: MAP2/tau repeat homology <MT2>
 F:245-276/Domain: MAP2/tau repeat homology <MT3>

Query Match 76.8%; Score 43; DB 2; Length 341;
 Best Local Similarity 77.8%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 9
 |||||
 DB 108 PPKTPPGSS 116

RESULT 5
 A28820
 microtubule-associated protein tau type 1 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
 C:Accession: A28820
 R:Lee, G.; Cowan, N.; Kirschner, M.
 Science 239, 285-288, 1988
 A:Title: The primary structure and heterogeneity of tau protein from mouse brain.
 A:Reference number: A94298; MUID:88099510; PMID:3122323
 A:Accession: A28820
 A:Molecule type: mRNA
 A:Residues: 1-364 <LEE>
 A:Cross-references: GB:M18776; NID:g201116; PIDN:AAA40166.1; PID:g201117
 C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
 C:Keywords: alternative splicing; microtubule binding; tandem repeat
 F:183-213/Domain: MAP2/tau repeat homology <MT1>
 F:214-244/Domain: MAP2/tau repeat homology <MT2>
 F:245-276/Domain: MAP2/tau repeat homology <MT3>

Query Match 76.8%; Score 43; DB 2; Length 364;
 Best Local Similarity 77.8%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 9
 |||||
 DB 108 PPKTPPGSS 116

RESULT 6
 S46264
 microtubule-associated protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 13-Aug-1999
 C:Accession: S46264
 R:Sadot, E.; Marx, R.; Barg, J.; Behar, L.; Ginzburg, I.
 J. Mol. Biol. 241, 325-331, 1994
 A:Title: Complete sequence of 3'-untranslated region of tau from rat central nervous system
 A:Reference number: S46264; MUID:94334997; PMID:8057376
 A:Accession: S46264
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-374 <SAD>
 A:Cross-references: EMBL:X79321; NID:g517393; PIDN:CAA55889.1; PID:g517394
 C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
 F:185-215/Domain: MAP2/tau repeat homology <MT1>
 F:216-246/Domain: MAP2/tau repeat homology <MT2>
 F:247-277/Domain: MAP2/tau repeat homology <MT3>
 F:278-309/Domain: MAP2/tau repeat homology <MT4>

Query Match 76.8%; Score 43; DB 2; Length 374;
 Best Local Similarity 77.8%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 9
 |||||
 DB 110 PPKTPPGSS 118

```

Db          422 PSPKTPPGS 430

RESULT 7
JS0306
microtubule-associated protein tau - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 31-Dec-1993
C:Accession: JS0306; A33574
R:Kosik, K.S.; Orecchio, L.D.; Bakalis, S.; Neve, R.L.
Neuron 2, 1389-1397, 1989
A:Title: Developmentally regulated expression of specific tau sequences.
A:Reference number: JS0306; MUID:90180457; PMID:2560640
A:Accession: JS0306
A:Molecule type: mRNA
A:Residues: 1-432 <KOS>
A>Note: the sequence shown is from adult rat brain
A>Note: the partial sequence from fetal rat brain is lacking residues 266-296; the fetal
A>Note: both fetal and adult forms were found in the paired helical filaments characteri
R:Kanai, Y.; Takemura, R.; Oshima, T.; Mori, H.; Ihara, Y.; Yanagisawa, M.; Masaki, T.;
J. Cell Biol. 109, 1173-1184, 1989
A:Title: Expression of multiple tau isoforms and microtubule bundle formation in fibrobl
A:Reference number: A33574; MUID:89359509; PMID:2504728
A:Accession: A33574
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-432 <KAN>
A>Note: a variant lacking residues 63-91 was also found
C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C:Keywords: alternative splicing; Alzheimer's disease; calmodulin binding; microtubule b
F:243-273/Domain: MAP2/tau repeat homology <MT1>
F:274-304/Domain: MAP2/tau repeat homology <MT1>
F:305-335/Domain: MAP2/tau repeat homology <MT2>
F:336-367/Domain: MAP2/tau repeat homology <MT3>
F:368-399/Domain: MAP2/tau repeat homology <MT4>
F:400-432/Domain: MAP2/tau repeat homology <MT4>
F:433-464/Domain: MAP2/tau repeat homology <MT4>
F:465-496/Domain: MAP2/tau repeat homology <MT4>
F:497-527/Domain: MAP2/tau repeat homology <MT4>
F:528-558/Domain: MAP2/tau repeat homology <MT4>
F:559-589/Domain: MAP2/tau repeat homology <MT4>
F:590-621/Domain: MAP2/tau repeat homology <MT4>

Query Match          76.8%; Score 43; DB 2; Length 432;
Best Local Similarity 77.8%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
| | | | |
Db 168 PSPKTPPGS 176

RESULT 8
A38235
microtubule-associated protein, 110K tau - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A38235
R:Goedert, M.; Spillantini, M.G.; Crowther, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 1983-1987, 1992
A:Title: Cloning of a big tau microtubule-associated protein characteristic of the perip
A:Reference number: A38235; MUID:92179305; PMID:1542696
A:Accession: A38235
A:Molecule type: mRNA
A:Residues: 1-686 <GOE>
A:Cross-references: GB:M84156; NID:9207157; PIDN:AAA42204.1; PID:9207158
A>Note: sequence extracted from NCBI backbone (NCBIN:87358, NCBIP:87359)
C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C:Keywords: alternative splicing; microtubule binding; tandem repeat
F:497-527/Domain: MAP2/tau repeat homology <MT1>
F:528-558/Domain: MAP2/tau repeat homology <MT2>
F:559-589/Domain: MAP2/tau repeat homology <MT3>
F:590-621/Domain: MAP2/tau repeat homology <MT4>

Query Match          76.8%; Score 43; DB 2; Length 686;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
| | | | |
Db 168 PSPKTPPGS 176

RESULT 9
A45301
microtubule-associated protein tau - mouse
N:Alternate names: microtubule binding protein tau
C:Species: Mus musculus (house mouse)
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 13-Aug-1999
C:Accession: A45301; S31658
R:Couchie, D.; Mavilia, C.; Georgieff, I.S.; Liem, R.K.; Shelanski, M.L.; Nunez, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 4378-4381, 1992
A:Title: Primary structure of high molecular weight tau present in the peripheral nervou
A:Reference number: A45301; MUID:92262443; PMID:1374898
A:Accession: A45301
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-733 <COU>
A>Note: this sequence is inconsistent with the nucleotide translation
R:Kenner, L.; Forsner, M.; Hutter, H.; Hoefler, G.; Kurzbaue, R.; Zatloukal, K.; Krisp
submitted to the EMBL Data Library, May 1992
A:Description: First observation of mRNA for a tau-protein from murine liver and kidney.
A:Reference number: S31658
A:Accession: S31658
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-733 <KEN>
C:Cross-references: EMBL:Z12133; NID:954263; PIDN:CAA78121.1; PID:9388534
C:Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C:Keywords: microtubule binding; tandem repeat
F:544-574/Domain: MAP2/tau repeat homology <MT1>
F:575-605/Domain: MAP2/tau repeat homology <MT2>
F:606-636/Domain: MAP2/tau repeat homology <MT3>
F:637-668/Domain: MAP2/tau repeat homology <MT4>

Query Match          76.8%; Score 43; DB 2; Length 733;
Best Local Similarity 77.8%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
| | | | |
Db 469 PSPKTPPGS 477

RESULT 10
S27771
RNA-directed DNA polymerase (EC 2.7.7.49) - African malaria mosquito transposon RT1 (fra
N:Alternate names: reverse transcriptase
C:Species: Anopheles gambiae (African malaria mosquito)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27771
R:Besansky, N.J.; Paskewitz, S.M.; Mills-Hamm, D.M.; Collins, F.H.
submitted to the EMBL Data Library, June 1992
A:Description: Distinct families of site-specific retrotransposons occupy identical positions
A:Reference number: S27770
A:Accession: S27770
A:Molecule type: DNA
A:Residues: 1-1212 <BES>
A:Cross-references: EMBL:M93690; NID:9159615; PID:9159617
C:Keywords: nucleotidyltransferase

Query Match          76.8%; Score 43; DB 2; Length 1212;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
| | | | |
Db 545 PKPGKPGGSN 554

RESULT 11
F96575

```

hypothetical protein F22G10.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F96575
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96787
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-153 <STO>
 A:Cross-references: GB:AE005173; NID:G10645344; PIDN:AAG21464.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F22G10.10
 A:Map position: 1

Query Match 75.0%; Score 42; DB 2; Length 153;
 Best Local Similarity 77.8%; Pred. No. 28;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGS 9
 |||||
 Db 43 PQSPSPGS 51

RESULT 12

G65084
 hypothetical protein b2985 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 13-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: G65084
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: G65084
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-237 <BLAT>
 A:Cross-references: GB:AE000381; GB:U00096; NID:G2367181; PIDN:AAC76021.1; PID:G1789358;
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: conserved hypothetical protein b2986

Query Match 75.0%; Score 42; DB 2; Length 237;
 Best Local Similarity 87.5%; Pred. No. 42;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KPSTPPGS 9
 |||||
 Db 84 KPSTPPGN 91

RESULT 13

D96787
 protein T4012.3 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 24-Aug-2001
 C:Accession: D96787
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96787
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-330 <STO>
 A:Cross-references: GB:AE005173; NID:G6721098; PIDN:AAF26752.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T4012.3
 A:Map position: 1
 C:Superfamily: thaumatin I

Query Match 75.0%; Score 42; DB 2; Length 330;
 Best Local Similarity 70.0%; Pred. No. 58;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 264 PKPTTPTGTS 273

RESULT 14

D71460
 probable membrane thiol proteinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: D71460
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: D71460
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-418 <ARN>
 A:Cross-references: GB:AE001360; GB:AE001273; NID:G3329342; PIDN:AAC68466.1; PID:G332934
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: CT868

Query Match 75.0%; Score 42; DB 2; Length 418;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPP 7
 |||||
 Db 128 PKPSTPP 134

RESULT 15

T31243
 hypothetical protein 915 - Sphingomonas aromaticivorans plasmid pNL1
 C:Species: Sphingomonas aromaticivorans
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T31243
 R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; G
 submitted to the EMBL Data Library, July 1998
 A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati
 A:Reference number: Z20992
 A:Accession: T31243
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-173 <ROM>
 A:Cross-references: EMBL:AF09317; NID:G3378261; PID:G3378383; PIDN:AAD03966.1
 C:Genetics:
 A:Gene: plasmid pNL1
 A:Note: orf915
 C:Superfamily: Sphingomonas aromaticivorans hypothetical protein 915

Query Match 73.2%; Score 41; DB 2; Length 173;

Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KPSTPPGS 9
| | | | |
Db 13 KPSSPPGS 20

Search completed: April 20, 2004, 10:27:42
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:20:04 ; Search time 40.4 Seconds
(without alignments)
69.938 Million cell updates/sec

Title: US-08-930-480A-7

Perfect score: 56

Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	10	2	Aaw09324 Peptide 1
2	56	100.0	10	5	Abg97610 Apolipop
3	56	100.0	16	2	Aaw71022 Mus muscu
4	56	100.0	17	5	Abg94379 N-termina
5	56	100.0	17	5	Abg80675 N-termina
6	56	100.0	17	6	Abg56408 Peptide 1
7	56	100.0	17	6	Abg44509 Peptide 1
8	56	100.0	17	6	Abg44509 Peptide 1
9	56	100.0	18	5	Abg94380 C-termina
10	56	100.0	18	5	Abg80676 C-termina
11	56	100.0	18	6	Abg56409 Peptide 1
12	56	100.0	18	6	Abg44510 Peptide 1
13	56	100.0	18	7	Abg44510 Peptide 1
14	56	100.0	42	3	Aay91034 Lipid-tag
15	56	100.0	46	2	Aar39337 Intercala
16	56	100.0	47	2	Aaw22021 Di-alpha-
17	56	100.0	50	2	Aar39338 Intercala
18	56	100.0	53	2	Aar39340 Intercala
19	56	100.0	53	2	Aar39339 Intercala
20	56	100.0	81	5	Abp51694 Plasmid p
21	56	100.0	81	5	Abb79463 Recombina
22	56	100.0	134	5	Abg94335 Human MIF
23	56	100.0	134	5	Abg94336 met-human
24	56	100.0	134	5	Abg80648 Human MIF
25	56	100.0	135	5	Abg94330 rMIF-C3 p

26	56	100.0	135	5	Abg94334 Human MIF
27	56	100.0	135	5	Abg80642 Rat MIF
28	56	100.0	135	5	Abg80647 Human MIF
29	56	100.0	136	5	Abg94348 Mouse C-I
30	56	100.0	136	5	Abg80660 Mouse II-
31	56	100.0	138	5	Abg94351 Human C-I
32	56	100.0	138	5	Abg80663 Human Iun
33	56	100.0	263	5	Abg80714 Human IgG
34	56	100.0	322	4	Abg20440 Antibody
35	56	100.0	325	4	Abg20438 Anti-FIX/
36	56	100.0	450	3	Aay44991 M79scFv-i
37	56	100.0	466	6	Abu62399 Chimeric
38	56	100.0	467	6	Abg38408 Mouse vir
39	56	100.0	531	2	Aar98007 PelB sign
40	45	80.4	782	4	Aar98007 Human pol
41	43	76.8	116	4	Aau48146 Propionib
42	43	76.8	116	6	Abm44665 Propionib
43	43	76.8	364	5	Abm57300 Mouse iac
44	43	76.8	374	6	Abm04837 Rat tau m
45	43	76.8	446	4	Abg06313 Novel hum

ALIGNMENTS

RESULT 1
AAW09324
ID AAW09324 standard; peptide; 10 AA.

AC AAW09324;

DT 10-JUN-1997 (first entry)

DE Peptide linker arm #2.

KW Chimeric; bispecific; DNA binding domain; trans; activator; repressor;
KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;
KW hyperproliferation; p53; tumour; oligomerisation.

OS Synthetic.

FN WO9630512-A1.

PD 03-OCT-1996.

PF 29-MAR-1996; 96WO-FR000477.

PR 31-MAR-1995; 95FR-00003841.

PA (RHON) RHONE POULENC RORER SA.

PI Bracco L, Schweighoffer F, Tocque B;

DR WPI; 1996-455359/45.

DR N-PSDB; AAT47997.

PT Conditional gene expression system triggered by e.g. infection or hyper-
proliferation - comprises novel bi-specific proteins having DNA-binding
domain and second domain specific for trans-activator or repressor, for
gene therapy.

PS Claim 23; Page 46; 81pp; French.

CC The invention relates to novel chimeric, bispecific proteins which
comprise: (a) a DNA binding domain and (b) a domain which binds a trans-
activator (TA), trans-repressor (TR) or their complexes, which are
characteristic of a physiological or physiopathological state. The novel
chimeric, bispecific proteins allow expression of a therapeutic protein
(e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
antibodies) to be regulated in response to particular conditions.
Examples include making the protein responsive to the presence of
particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7)

CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
 CC will be expressed in those cells infected by that pathogen. Similarly,
 CC where the chimeric protein responds to a cellular protein typical of a
 CC hyperproliferative state (esp. wild-type and mutant p53), expression can
 CC be restricted to tumour cells. The sequence presented here is an example
 CC of a peptide linker "arm" which connects the DNA binding domain to the TA
 CC binding domain
 CC
 CC Sequence 10 AA;

Query Match 100.0%; Score 56; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 1 PKPSTPPGSS 10

RESULT 2
 ABG97610
 ID ABG97610 standard; peptide; 10 AA.
 XX AC ABG97610;
 XX 17-DEC-2002 (first entry)
 XX DE Apolipoprotein analogue 1 (Apo A1) associated spacer peptide #4.
 XX KW Apolipoprotein analogue; Apo A; arteriosclerosis; endotoxin removal;
 XX KW angina pectoris; myocardial infarction; arterial stenosis; claudication;
 XX KW carotid stenosis; cerebral arterial stenosis; gene therapy; cholesterol;
 XX KW cardiovascular disease; spacer peptide.

XX Mus sp.
 XX WO200238609-A2.
 XX 16-MAY-2002.
 XX 09-NOV-2001; 2001WO-DK000739.
 XX 10-NOV-2000; 2000DK-00001682.
 XX 15-JAN-2001; 2001DK-00000057.
 XX 26-JAN-2001; 2001US-0264022P.
 XX (PROT-) PROTEOPHARMA APS.
 XX Graversen J, Moestrup S;
 XX WPI; 2002-527481/56.

XX Novel apolipoprotein construct comprising apolipoprotein A linked to
 PT carbohydrate, peptide or protein heterologous group, useful for treating
 PT plaque/unstable angina pectoris, myocardial infarction, arterial
 PT stenoses.

XX Claim 8; Page 54; 113pp; English.
 XX The invention describes an Apolipoprotein (Apo) construct (I) for use as
 CC medicament having general formula apo-A-X, where apo-A is an
 CC apolipoprotein component such as apolipoprotein A1, AII or AIV, or its
 CC analogue or variant, and X is heterologous group e.g., amino acid,
 CC peptide, protein, carbohydrate or a nucleic acid, providing that when (I)
 CC consists of exactly two identical, native apolipoproteins these are
 CC linked serially. (I) is useful for preparing a pharmaceutical composition
 CC which further comprises excipients, adjuvants, additives, such as
 CC phospholipids, cholesterol or triglycerides. (I) is useful for treating
 CC and/or preventing arteriosclerosis, for removing endotoxins, for treating
 CC angina pectoris including plaque or unstable angina pectoris, myocardial
 CC infarction, arterial stenoses such as claudication, carotid stenosis,
 CC cerebral arterial stenosis and other cardiovascular diseases. The nucleic
 CC acid (II) encoding (I) is useful for gene therapy, where the DNA sequence

CC encoding (I) is used for transfection or infection of at least one cell
 CC population comprising macrophages or liver cells. (I) has a half-life of
 CC at least the half-life of native apoA-I, A-II or A-IV, preferably two
 CC times higher or more preferably 10 times higher than the half-life of the
 CC apoA molecules. (I) also has a higher binding affinity to cholesterol
 CC compared to native apoA-I, A-II or A-IV. (I) causes substantially no
 CC immune response in humans. This is the amino acid sequence of a spacer
 CC peptide used to link human apolipoprotein (Apo) or an Apo analogue
 CC protein to a heterologous moiety
 CC
 CC Sequence 10 AA;

Query Match 100.0%; Score 56; DB 5; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 1 PKPSTPPGSS 10

RESULT 3
 AAW71022
 ID AAW71022 standard; protein; 16 AA.
 XX AC AAW71022;
 XX 09-NOV-1998 (first entry)
 XX DE Mus musculus Cgamma3 gene hinge region.
 XX KW Cgamma3 gene; transgenic; screening; therapeutic; nephritis;
 XX KW systemic lupus erythematosus.

XX Mus musculus.
 XX WO9837174-A1.
 XX 27-AUG-1998.
 XX 18-FEB-1998; 98WO-US003027.
 XX 20-FEB-1997; 97US-00803120.
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 XX Schreiber JR, Greenspan NS, Threadgill DS, Magnuson T;
 XX WPI; 1998-467546/40.
 XX N-PSDB; AAV43042.

XX New transgenic animals lacking IgG3 - is used for screening candidate
 PT therapeutic compounds, in particular for activity against bacterial
 PT infection or nephritis.
 XX Disclosure; Fig 1; 60pp; English.

XX The sequence is that of the hinge region of the protein encoded by the
 CC mouse Cgamma3 gene
 CC
 CC Sequence 16 AA;

Query Match 100.0%; Score 56; DB 2; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 4 PKPSTPPGSS 13

RESULT 4
 ABG94379

ID ABG94379 standard; peptide; 17 AA.
 XX ABG94379;
 AC
 DT 10-DEC-2002 (first entry)
 XX
 DE N terminal gamma 3 amino acid linker.
 XX
 KW Human; mouse; rat; antimicrobial; antiallergic; immunomodulatory;
 KW cytostatic; antiviral; antidiabetic; hypoglycaemic; antigen array;
 KW vaccine; infectious disease.
 XX
 OS Synthetic.
 XX
 PN WO200256905-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 21-JAN-2002; 2002WO-IB000166.
 XX
 PR 19-JAN-2001; 2001US-0262379P.
 PR 04-MAY-2001; 2001US-0288549P.
 PR 05-OCT-2001; 2001US-0326998P.
 PR 07-NOV-2001; 2001US-0331045P.
 XX
 PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
 XX
 PI Renner WA, Bachmann M, Tissot A, Maurer P, Lechner F, Sebbel P;
 PI Piossek C;
 XX
 DR WPI; 2002-627351/67.
 XX
 PT Molecular antigen array used in the production of vaccines for infectious
 PT diseases.
 XX
 PS Disclosure; Page 49; 441pp; English.
 XX
 CC This invention relates to a novel ordered and repetitive antigen array
 CC used in the production of vaccines for infectious diseases. The invention
 CC also discloses a composition comprising a non-natural molecular scaffold
 CC comprising a core particle selected from a non-natural molecular scaffold
 CC origin and a core particle of natural origin and an organiser comprising
 CC at least one first attachment site, where the organiser is connected to
 CC the core particle by at least one covalent bond. Also disclosed is an
 CC antigen or antigenic determinant with at least one second attachment
 CC site, where the antigen or antigenic determinant is amyloid beta peptide
 CC (Abeta1-42) or its fragment and where the second attachment site is
 CC selected from an attachment site not naturally occurring with the antigen
 CC or antigenic determinant and an attachment site naturally occurring with
 CC the antigen or antigenic determinant, where the second attachment site is
 CC capable of association through at least one non-peptide bond to the first
 CC attachment site and where the antigen or antigenic determinant and the
 CC scaffold interact through the association to form an ordered and
 CC repetitive antigen array. The invention also comprises a coat protein
 CC capable of forming a capsid which comprises mutant Qbeta coat proteins
 CC having an amino acid sequence selected from five amino acid sequences
 CC fully defined in the specification. The compounds of the invention may
 CC have antimicrobial, antiallergic, immunomodulatory, cytostatic,
 CC antiviral, antidiabetic, or hypoglycaemic activities and may be used in
 CC immunization and as a vaccine. The present sequence represents a peptide
 CC sequence used to create the compositions of the invention
 XX
 SQ Sequence 17 AA;

Query Match 100.0%; Score 56; DB 5; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PKPSTPPGSS 10
 | | | | |
 Db 4 PKPSTPPGSS 13

RESULT 5
 ABG80675
 ID ABG80675 standard; peptide; 17 AA.
 XX
 AC ABG80675;
 XX
 DT 29-NOV-2002 (first entry)
 XX
 DE N-terminal gamma 3 linker peptide.
 XX
 KW Molecular antigen array; vaccine; antigen; antimicrobial;
 KW molecular scaffold; amyloid beta; Abeta 1-42; influenza;
 KW graft versus host disease; IGE-mediated allergic reaction; anaphylaxis;
 KW adult respiratory distress syndrome; ARDS; Crohn's disease;
 KW allergic asthma; acute lymphoblastic leukaemia; non-Hodgkin's lymphoma;
 KW Grave's disease; systemic lupus erythematosus; osteoporosis;
 KW inflammatory immune disease; myasthenia gravis; multiple sclerosis;
 KW immunoproliferative disease lymphadenopathy; Alzheimer's disease;
 KW angioimmunoproliferative lymphadenopathy; immunoblastic lymphadenopathy;
 KW rheumatoid arthritis; diabetes; infectious disease; factor Xa;
 KW enterokinase; cysteine-containing linker.
 XX
 OS Synthetic.
 XX
 PN WO200256907-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 21-JAN-2002; 2002WO-IB000168.
 XX
 PR 19-JAN-2001; 2001US-0262379P.
 PR 04-MAY-2001; 2001US-0288549P.
 PR 05-OCT-2001; 2001US-0326998P.
 PR 07-NOV-2001; 2001US-0331045P.
 XX
 PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
 PA (NOVS) NOVARTIS PHARMA AG.
 PA (MAUR) MAURER P.
 PA (LECH) LECHNER F.
 PA (ORTM) ORTMANN R.
 PA (LUEO) LUEOEND R.
 PA (STAU) STAUFENBIEL M.
 PA (FREY) FREY P.
 XX
 PI Maurer P, Lechner F, Ortmann R, Lueoend R, Staufenbiel M, Frey P;
 PI Renner WA, Bachmann M, Tissot A, Sebbel P, Piossek C;
 XX
 DR WPI; 2002-636514/68.
 XX
 PT Molecular antigen array used in the production of vaccines for infectious
 PT diseases.
 XX
 PS Claim 35; Page 49; 418pp; English.
 XX
 CC The invention relates to a composition comprising: (a) a non-natural
 CC molecular scaffold comprising: (i) a core particle selected from: (1) a
 CC core particle of a non-natural origin; and (2) a core particle of natural
 CC origin; and (ii) an organiser comprising at least one first attachment
 CC site, where the organiser is connected to the core particle by at least
 CC one covalent bond; (b) an antigen or antigenic determinant with at least
 CC one second attachment site, where the antigen or antigenic determinant is
 CC amyloid beta peptide (Abeta 1-42) or its fragment, and where the second
 CC attachment site is selected from: (i) an attachment site not naturally
 CC occurring with the antigen or antigenic determinant; and (ii) an
 CC attachment site naturally occurring with the antigen or antigenic
 CC determinant, where the second attachment site is capable of association
 CC through at least one non-peptide bond to the first attachment site; and
 CC where the antigen or antigenic determinant and the scaffold interact
 CC through the association to form an ordered and repetitive antigen array.
 CC Also included is a process for producing a non-naturally occurring
 CC ordered and repetitive antigen array. The composition is used in
 CC immunisation and as a vaccine for diseases such as influenza, graft
 CC versus host disease, IGE-mediated allergic reactions, anaphylaxis, adult

respiratory distress syndrome (ARDS), Crohn's disease, allergic asthma, acute lymphoblastic leukaemia, non-Hodgkin's lymphoma, Grave's disease, systemic lupus erythematosus, inflammatory immune diseases, myasthenia gravis, immunoproliferative disease lymphadenopathy, angioimmunoproliferative lymphadenopathy, immunoblastic lymphadenopathy, rheumatoid arthritis, diabetes, multiple sclerosis, Alzheimer's disease, osteoporosis and infectious diseases. The antigens are modified to possess a cleavage site (enterokinase or factor Xa) and a Cysteine- containing N- or C-terminal linker peptide which serves as the attachment point to a virus like particle or bacterial protein (the scaffold protein). The present sequence is a cysteine-containing linker peptide used in the molecular antigen array

Sequence 17 AA;

Query Match 100.0%; Score 56; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
DB 4 PKPSTPPGSS 13
|||||

RESULT 6
ABR56408
ID ABR56408 standard; peptide; 17 AA.

AC ABR56408;

DT 28-JUL-2003 (first entry)

DE Peptide linker #8.

KW Antigen presenting cell; APC; immune response; virus like particle; VLP;
KW cytostatic; virucide; antibacterial; antiparasitic; fungicide;
KW antiallergic; immunosuppressive; antiaddictive; antiinflammatory;
KW antithyroid; antidiabetic; neuroprotective; nontropic; osteopathic;
KW antirheumatic; antiarthritic; vaccine; immunisation; infectious disease;
KW anti-viral protection; tumour; allergy; drug addition; Crohn's disease;
KW Graft-versus-host disease; Grave's disease; diabetes; multiple sclerosis;
KW Alzheimer's disease; osteoporosis; rheumatoid arthritis;
KW inflammatory autoimmune disease.

OS Synthetic.

PN WO2003024480-A2.

PD 27-MAR-2003.

PF 16-SEP-2002; 2002WO-IB004252.

PR 14-SEP-2001; 2001US-0318967P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

XX Bachmann MF, Storni T, Lechner F;

XX WPI; 2003-363095/34.

XX A composition, useful for enhancing an immune response against an antigen or a virus-like particle, enhancing anti-viral protection in an animal, or immunizing or treating tumors or infectious diseases, e.g. viral infections.

XX Disclosure; Page 65; 243pp; English.

XX The present invention describes a composition (C) for enhancing an immune response against an antigen or a virus-like particle in an animal. (C) comprises a virus-like particle (VLP) bound to at least one antigen, or a VLP capable of being recognised by the immune system of the animal. Also described: (1) enhancing an immune response against an antigen or a VLP in an animal comprising introducing (C) into the animal; (2) vaccines

CC comprising (C) together with a pharmaceutical diluent, carrier or excipient; (3) immunising or treating an animal comprising administering the vaccine to the animal, or priming or boosting a T cell response in the animal by administering the vaccine; and (4) enhancing anti-viral protection in an animal comprising introducing (C) into the animal. (C) has cytostatic, virucide, antibacterial, antiparasitic, fungicide, antiallergic, immunosuppressive, antiaddictive, antiinflammatory, antithyroid, antidiabetic, neuroprotective, nontropic, osteopathic, antirheumatic and antiarthritic activities. (C) or the vaccines can be used for enhancing an immune response against an antigen or a VLP in an animal, enhancing anti-viral protection in an animal, or immunising or treating tumors and infectious diseases such as viral, bacterial, parasitic or fungal infections. The vaccine compositions are also useful for preventing or treating allergies, drug addiction, multiple sclerosis, disease, Crohn's disease, Grave's disease, diabetes, rheumatoid arthritis, Alzheimer's disease, osteoporosis, rheumatoid arthritis, or inflammatory autoimmune disease. ACC69838 to ACC69852 and ABR56401 to ABR56599 CC represent sequences used in the exemplification of the present invention

XX Sequence 17 AA;

Query Match 100.0%; Score 56; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10
DB 4 PKPSTPPGSS 13
|||||

RESULT 7
ABR44509
ID ABR44509 standard; peptide; 17 AA.

AC ABR44509;

DT 25-JUL-2003 (first entry)

DE Peptide linker #8.

KW Immunostimulatory; virus-like particle; bacteriophage; HSV; LCMV;
KW hepatitis B virus; lymphocytic choriomeningitis virus; vaccine;
KW immunostimulant; cytostatic; antiallergic; virucide; antibacterial;
KW immune response; immunisation; allergy; tumour; breast cancer;
KW neuroblastoma; leukaemia; viral disease; influenza; hepatitis; measles;
KW chicken pox; bacterial infection; tuberculosis; pneumonia; syphilis.

XX Synthetic.

XX WO2003024481-A2.

XX 27-MAR-2003.

XX 16-SEP-2002; 2002WO-IB004132.

XX 14-SEP-2001; 2001US-0318994P.

XX 22-APR-2002; 2002US-0374145P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

XX (MAUR/) MAURER P.

XX (TISS/) TISSOT A.

XX (SCHW/) SCHWARZ K.

XX (MEIJ/) MEIJERINK E.

XX (LIPO/) LIPOWSKY G.

XX (PUMP/) PUMPENS P.

XX (CIEL/) CIELENS I.

XX (RENH/) RENHOFFA R.

XX Maurer P, Tissot A, Schwarz K, Meijerink E, Lipowsky G;
PI Pumpens P, Cielens I, Renhofs R, Bachmann MF, Storni T;
XX WPI; 2003-354564/33.

PT New compositions comprising immunostimulatory substances packaged into
 PT virus-like particles, useful as a vaccine for enhancing an immune
 PT response in animals, e.g. for treating or preventing allergies, tumors or
 PT viral infections.

PS Disclosure; Page 75; 322pp; English.

XX The present invention describes a composition (C) for enhancing an immune
 CC response in an animal. (C) comprises a virus-like particle (VLP), and an
 CC immunostimulatory substance. The immunostimulatory substance is bound to
 CC the VLP. Also described: (1) enhancing an immune response in an animal by
 CC introducing (C) into the animal; (2) producing (C) for enhancing an
 CC immune response in an animal; (3) vaccines comprising (C) together with a
 CC pharmaceutical diluent, carrier or excipient; and (4) immunising or
 CC treating an animal by: (a) administering the vaccine to the animal; (b)
 CC priming a T cell response in the animal by administering the vaccine; or
 CC (c) boosting a T cell response in the animal by administering the
 CC vaccine. (C) has immunostimulant, cytostatic, antiallergic, virucide and
 CC antibacterial activities. (i) can be used in vaccines for enhancing an
 CC immune response in an animal, particularly a mammal or human.

CC Specifically, (C) is useful for enhancing a B cell response, a T cell
 CC response, or a cytotoxic T-lymphocyte (CTL) response. (C) or a vaccine
 CC comprising (C) can also be used for immunising or treating an animal,
 CC e.g. humans, sheep, horses, cattle, pigs, dogs, cats, rats, birds,
 CC reptiles or fish. (C) is particularly useful in prophylactic or
 CC therapeutic vaccines against allergies, tumours (e.g. breast cancers,
 CC neuroblastoma, or leukaemia), viral diseases (e.g. influenza, hepatitis,
 CC measles or chicken pox), or bacterial infections (e.g. tuberculosis,
 CC pneumonia or syphilis). ACC69790 to ACC69815 and ABR44502 to ABR44612
 CC represent sequences used in the exemplification of the present invention

XX Sequence 17 AA;

Query Match 100.0%; Score 56; DB 6; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
 |||||
 Db 4 PKPSTPPGSS 13

RESULT 8

ADD24209
 ID ADD24209 standard; peptide; 17 AA.

XX AC ADD24209;

DT 15-JAN-2004 (first entry)

XX Linker peptide 3 related to prion disease vaccines.

XX vaccine composition; virus-like particle; core particle;
 KW first attachment site; antigen; antigenic determinant; prion protein;
 KW PrP; PrP peptide; vaccine; neuroprotective; antiinflammatory;
 KW prion disease; Bovine Spongiform Encephalopathy; BSE;
 KW Creutzfeldt-Jakob Disease; linker peptide.

XX Synthetic.

XX WO20003059386-A2.

XX PD 24-JUL-2003.

XX PF 17-JAN-2003; 2003WO-EP000460.

XX PR 18-JAN-2002; 2002US-00050902.

XX PR 21-JAN-2002; 2002WO-IB000166.

XX PR 08-JUL-2002; 2002US-0393725P.

XX PR 18-JUL-2002; 2002US-0396590P.

XX PA (CYTO-) CYTOS BIOTECHNOLOGY AG.

PI Bachmann M, Maurer P, Pelliccioli E, Renner WA;
 DR WPI; 2003-598483/56.

XX A vaccine composition for preventing or treating prion diseases (e.g.
 PT Creutzfeldt-Jakob Disease) comprises a virus-like particle (e.g. RNA-
 PT phage) and at least one prion protein or peptide bound to the virus-like
 PT particle.

XX Disclosure; Page 59; 246pp; English.

XX This invention relates to a novel vaccine composition comprising a virus-
 CC like or a core particle with at least one first attachment site and at
 CC least one antigen or antigenic determinant that is a prion protein (PrP)
 CC or its dimer, or a PrP peptide, the antigen or antigenic determinant
 CC being bound to the virus-like or core particle. The vaccine of the
 CC invention may have neuroprotective or antiinflammatory activity. The
 CC composition is useful as a medicament or in manufacturing a medicament
 CC for the treatment or prevention of prion diseases. The prion diseases may
 CC include Bovine Spongiform Encephalopathy (BSE) or Creutzfeldt-Jakob
 CC Disease. The present sequence is that of a linker peptide which is
 CC related to the invention.

XX Sequence 17 AA;

Query Match 100.0%; Score 56; DB 7; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
 |||||
 Db 4 PKPSTPPGSS 13

RESULT 9

ABG94380
 ID ABG94380 standard; peptide; 18 AA.

XX AC ABG94380;

XX 10-DEC-2002 (first entry)

XX C terminal gamma 3 amino acid linker.

XX Human; mouse; rat; antimicrobial; antiallergic; immunomodulatory;
 KW Cytostatic; antiviral; antidiabetic; hypoglycaemic; antigen array;
 KW vaccine; infectious disease.

XX Synthetic.

XX WO200256905-A2.

XX PD 25-JUL-2002.

XX PF 21-JAN-2002; 2002WO-IB000166.

XX PR 19-JAN-2001; 2001US-0262379P.

XX PR 04-MAY-2001; 2001US-0288549P.

XX PR 05-OCT-2001; 2001US-0326998P.

XX PR 07-NOV-2001; 2001US-0331045P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

XX Renner WA, Bachmann M, Tissot A, Maurer P, Lechner F, Sebbel P;

PI Piossek C;

XX WPI; 2002-627351/67.

XX Molecular antigen array used in the production of vaccines for infectious
 PT diseases.

XX Disclosure; Page 49; 441pp; English.

CC This invention relates to a novel ordered and repetitive antigen array
 CC used in the production of vaccines for infectious diseases. The invention
 CC also discloses a composition comprising a non-natural molecular scaffold
 CC comprising a core particle selected from a core particle of a non-natural
 CC origin and a core particle of natural origin and an organiser comprising
 CC at least one first attachment site, where the organiser is connected to
 CC the core particle by at least one covalent bond. Also disclosed is an
 CC antigen or antigenic determinant with at least one second attachment
 CC site, where the antigen or antigenic determinant is amyloid beta peptide
 CC (Abeta1-42) or its fragment and where the second attachment site is
 CC selected from an attachment site not naturally occurring with the antigen
 CC or antigenic determinant and an attachment site naturally occurring with
 CC the antigen or antigenic determinant, where the second attachment site is
 CC capable of association through at least one non-peptide bond to the first
 CC attachment site and where the antigen or antigenic determinant and the
 CC scaffold interact through the association to form an ordered and
 CC repetitive antigen array. The invention also comprises a coat protein
 CC capable of forming a capsid which comprises mutant beta coat proteins
 CC having an amino acid sequence selected from five amino acid sequences
 CC fully defined in the specification. The compounds of the invention may
 CC have antimicrobial, antiallergic, immunomodulatory, cytostatic,
 CC antiviral, antidiabetic, or hypoglycaemic activities and may be used in
 CC immunization and as a vaccine. The present sequence represents a peptide
 CC sequence used to create the compositions of the invention

XX Sequence 18 AA;

SQ Query Match 100.0%; Score 56; DB 5; Length 18;
 Best Local Similarity 100.0%; Pred. No. 1.5; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 1 PKPSTPPGSS 10

RESULT 10

ABG80676
 ID ABG80676 standard; peptide; 18 AA.

AC ABG80676;

XX 29-NOV-2002 (first entry)

DE C-terminal gamma 3 linker peptide.

XX Molecular antigen array; vaccine; antigen; antimicrobial;
 KW molecular scaffold; amyloid beta; Abeta 1-42; influenza;
 KW graft versus host disease; IGE-mediated allergic reaction; anaphylaxis;
 KW adult respiratory distress syndrome; ARDS; Crohn's disease;
 KW allergic asthma; acute lymphoblastic leukaemia; non-Hodgkin's lymphoma;
 KW Grave's disease; systemic lupus erythematosus; osteoporosis;
 KW inflammatory immune disease; myasthenia gravis; multiple sclerosis;
 KW immunoproliferative disease lymphadenopathy; Alzheimer's disease;
 KW angioimmunoproliferative lymphadenopathy; immunoblastic lymphadenopathy;
 KW rheumatoid arthritis; diabetes; infectious disease; factor Xa;
 KW enterokinase; cysteine-containing linker.

OS Synthetic.

XX WO200256907-A2.

PN 25-JUL-2002.

XX 21-JAN-2002; 2002WO-IB000168.

XX 19-JAN-2001; 2001US-0262379P.

PR 04-MAY-2001; 2001US-0288549P.

PR 05-OCT-2001; 2001US-0326998P.

PR 07-NOV-2001; 2001US-0331045P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

PA (NOVS) NOVARTIS PHARMA AG.

PA (MAUR/) MAURER P.
 PA (LECH/) LECHNER F.
 PA (ORTM/) ORTMANN R.
 PA (LUEO/) LUEOEND R.
 PA (STAU/) STAUFENBIEL M.
 PA (FREY/) FREY P.

XX Maurer P, Lechner F, Ortmann R, Lueoend R, Staufenbiel M, Frey P;
 PI Renner WA, Bachmann M, Tissot A, Sebbel P, Piossek C;

XX WPI; 2002-636514/58.

XX Molecular antigen array used in the production of vaccines for infectious
 PT diseases.

PS Claim 35; Page 49; 418pp; English.

XX The invention relates to a composition comprising: (a) a non-natural
 CC molecular scaffold comprising: (i) a core particle selected from: (1) a
 CC core particle of a non-natural origin; and (2) a core particle of natural
 CC origin; and (ii) an organiser comprising at least one first attachment
 CC site, where the organiser is connected to the core particle by at least
 CC one covalent bond; (b) an antigen or antigenic determinant with at least
 CC one second attachment site, where the antigen or antigenic determinant is
 CC amyloid beta peptide (Abeta 1-42) or its fragment, and where the second
 CC attachment site is selected from: (i) an attachment site not naturally
 CC occurring with the antigen or antigenic determinant; and (ii) an
 CC attachment site naturally occurring with the antigen or antigenic
 CC determinant where the second attachment site is capable of association
 CC through at least one non-peptide bond to the first attachment site; and
 CC where the antigen or antigenic determinant and the scaffold interact
 CC through the association to form an ordered and repetitive antigen array.
 CC Also included is a process for producing a non-naturally occurring
 CC ordered and repetitive antigen array. The composition is used in
 CC immunisation and as a vaccine for diseases such as influenza, graft
 CC versus host disease, IGE-mediated allergic reactions, anaphylaxis, adult
 CC respiratory distress syndrome (ARDS), Crohn's disease, allergic asthma,
 CC acute lymphoblastic leukaemia, non-Hodgkin's lymphoma, Grave's disease,
 CC systemic lupus erythematosus, inflammatory immune diseases, myasthenia
 CC gravis, immunoproliferative disease lymphadenopathy,
 CC angioimmunoproliferative lymphadenopathy, immunoblastic lymphadenopathy,
 CC rheumatoid arthritis, diabetes, multiple sclerosis, Alzheimer's disease,
 CC osteoporosis and infectious diseases. The antigens are modified to possess
 CC a cleavage site (enterokinase or factor Xa) and a Cysteine- containing N-
 CC or C-terminal linker peptide which serves as the attachment point to a
 CC virus like particle or bacterial protein (the scaffold protein). The
 CC present sequence is a cysteine-containing linker peptide used in the
 CC molecular antigen array

XX Sequence 18 AA;

SQ Query Match 100.0%; Score 56; DB 5; Length 18;
 Best Local Similarity 100.0%; Pred. No. 1.5; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0;

QY 1 PKPSTPPGSS 10
 |||||
 Db 1 PKPSTPPGSS 10

RESULT 11

ABR56409
 ID ABR56409 standard; peptide; 18 AA.

XX ABR56409;

XX 28-JUL-2003 (first entry)

DT Peptide linker #9.

XX Antigen presenting cell; APC; immune response; virus like particle; VLP;
 DE cytostatic; virucide; antibacterial; antiparasitic; fungicide;
 XX antiallergic; immunosuppressive; antiaddictive; antiinflammatory;

KW antihypoid; antidiabetic; neuroprotective; nootropic; osteopathic;
 KW antirheumatic; antiarthritic; vaccine; immunisation; infectious disease;
 KW anti-viral protection; tumour; allergy; drug addiction; Crohn's disease;
 KW graft-versus-host disease; Grave's disease; diabetes; multiple sclerosis;
 KW Alzheimer's disease; osteoporosis; rheumatoid arthritis;
 KW inflammatory autoimmune disease.

XX Synthetic.

PN WO2003024480-A2.

XX 27-MAR-2003.

PF 16-SEP-2002; 2002WO-IB004252.

PR 14-SEP-2001; 2001US-0318967P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

XX Bachmann MF, Storni T, Lechner F;

DR WPI; 2003-363095/34.

XX A composition, useful for enhancing an immune response against an antigen
 PT or a virus-like particle, enhancing anti-viral protection in an animal,
 PT or immunizing or treating tumors or infectious diseases, e.g. viral
 PT infections.

PS Disclosure; Page 65; 243pp; English.

XX The present invention describes a composition (C) for enhancing an immune
 CC response against an antigen or a virus-like particle in an animal. (C)
 CC comprises a virus-like particle (VLP) bound to at least one antigen, or a
 CC VLP capable of being recognised by the immune system of the animal. Also
 CC described: (1) enhancing an immune response against an antigen or a VLP
 CC in an animal comprising introducing (C) into the animal; (2) vaccines
 CC comprising (C) together with a pharmaceutical diluent, carrier or
 CC excipient; (3) immunising or treating an animal comprising administering
 CC the vaccine to the animal, or priming or boosting a T cell response in
 CC the animal by administering the vaccine; and (4) enhancing anti-viral
 CC protection in an animal comprising introducing (C) into the animal. (C)
 CC has cytostatic, virucide, antibacterial, antiparasitic, fungicide,
 CC antiallergic, immunosuppressive, antiaddictive, antiinflammatory,
 CC antithyroid, antidiabetic, neuroprotective, nootropic, osteopathic,
 CC antirheumatic and antiarthritic activities. (C) or the vaccines can be
 CC used for enhancing an immune response against an antigen or a VLP in an
 CC animal, enhancing anti-viral protection in an animal, or immunising or
 CC treating tumors and infectious diseases such as viral, bacterial,
 CC parasitic or fungal infections. The vaccine compositions are also useful
 CC for preventing or treating allergies, drug addiction, graft-versus-host
 CC disease, Crohn's disease, Grave's disease, diabetes, multiple sclerosis,
 CC Alzheimer's disease, osteoporosis, rheumatoid arthritis, or inflammatory
 CC autoimmune disease. ACC69838 to ACC69852 and ABR56401 to ABR56599
 CC represent sequences used in the exemplification of the present invention

XX Sequence 18 AA;

Query Match 100.0%; Score 56; DB 6; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPSTPPGSS 10

Db 1 PKPSTPPGSS 10

RESULT 12

ABR44510

ID ABR44510 standard; peptide; 18 AA.

AC ABR44510;

XX 25-JUL-2003 (first entry)

XX

DB

XX

KW

KW

KW

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PA

PA

PA

PA

PA

PA

PA

PI

PI

XX

XX

DR

XX

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

XX

Peptide linker #9.

Immunostimulatory; virus-like particle; bacteriophage; HBV; LCMV;

Hepatitis B virus; lymphocytic choriomeningitis virus; vaccine;

immunostimulant; cytostatic; antiallergic; virucide; antibacterial;

immune response; immunisation; allergy; tumour; breast cancer;

neuroblastoma; leukaemia; viral disease; influenza; hepatitis; measles;

chicken pox; bacterial infection; tuberculosis; pneumonia; syphilis.

XX Synthetic.

WO2003024481-A2.

XX 27-MAR-2003.

PF 16-SEP-2002; 2002WO-IB004132.

PR 14-SEP-2001; 2001US-0318994P.

PR 22-APR-2002; 2002US-0374145P.

XX (CYTO-) CYTOS BIOTECHNOLOGY AG.

PA (MAUR/) MAURER P.

PA (TISS/) TISSOT A.

PA (SCHW/) SCHWARZ K.

PA (MEIJ/) MEIJERINK E.

PA (LIPO/) LIPOWSKY G.

PA (PUMP/) PUMPENS P.

PA (CIEL/) CIELENS I.

PA (RENH/) RENHOFFA R.

XX

PI

PI

XX

XX

DR

XX

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

XX

XX

XX

Query Match 100.0%; Score 56; DB 6; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 18 AA;

New compositions comprising immunostimulatory substances packaged into
 virus-like particles, useful as a vaccine for enhancing an immune
 response in animals, e.g. for treating or preventing allergies, tumors or
 viral infections.

Disclosure; Page 75; 322pp; English.

The present invention describes a composition (C) for enhancing an immune
 response in an animal. (C) comprises a virus-like particle (VLP), and an
 immunostimulatory substance. The immunostimulatory substance is bound to
 the VLP. Also described: (1) enhancing an immune response in an animal by
 introducing (C) into the animal; (2) producing (C) for enhancing an
 immune response in an animal; (3) vaccines comprising (C) together with a
 pharmaceutical diluent, carrier or excipient; and (4) immunising or
 treating an animal by: (a) administering the vaccine to the animal; (b)
 priming a T cell response in the animal by administering the vaccine; or
 (c) boosting a T cell response in the animal by administering the vaccine and
 vaccine. (C) has immunostimulant, cytostatic, antiallergic, virucide and
 antibacterial activities. (I) can be used in vaccines for enhancing an
 immune response in an animal, particularly a mammal or human.
 Specifically, (C) is useful for enhancing a B cell response, a T cell
 response, or a cytotoxic T-lymphocyte (CTL) response. (C) or a vaccine
 comprising (C) can also be used for immunising or treating an animal,
 e.g. humans, sheep, horses, cattle, pigs, dogs, cats, rats, birds,
 reptiles or fish. (C) is particularly useful in prophylactic or
 therapeutic vaccines against allergies, tumours (e.g. breast cancers,
 neuroblastoma, or leukaemia), viral diseases (e.g. influenza, hepatitis,
 measles or chicken pox), or bacterial infections (e.g. tuberculosis,
 pneumonia or syphilis). ACC69790 to ACC69815 and ABR44502 to ABR44512
 represent sequences used in the exemplification of the present invention

QY 1 PKPSTPPGSS 10
AC |||||
DB 1 PKPSTPPGSS 10

RESULT 13

ADD24210
ID ADD24210 standard; peptide; 18 AA.

XX AC
XX ADD24210;

XX DT 15-JAN-2004 (first entry)

XX DE Linker peptide 4 related to prion disease vaccines.

XX KW vaccine composition; virus-like particle; core particle;
KW first attachment site; antigen; antigenic determinant; prion protein;
KW PrP; PrP peptide; vaccine; neuroprotective; antiinflammatory;
KW prion disease; Bovine Spongiform Encephalopathy; BSE;
KW Creutzfeldt-Jakob Disease; linker peptide.

XX OS Synthetic.

XX PN WO2003059386-A2.

XX PD 24-JUL-2003.

XX PF 17-JAN-2003; 2003WO-EP000460.

XX PR 18-JAN-2002; 2002US-00050902.

XX PR 21-JAN-2002; 2002WO-18000166.

XX PR 08-JUL-2002; 2002US-0393725P.

XX PR 18-JUL-2002; 2002US-0396590P.

XX PA (CYTO-) CYTOS BIOTECHNOLOGY AG.

XX PI Bachmann M, Maurer P, Pelliccioli E, Renner WA;

XX DR WPI; 2003-598483/56.

XX PT A vaccine composition for preventing or treating prion diseases (e.g.
PT Creutzfeldt-Jakob Disease) comprises a virus-like particle (e.g. RNA-
PT phage), and at least one prion protein or peptide bound to the virus-like
PT particle.

XX PS Disclosure; Page 59; 246pp; English.

XX CC This invention relates to a novel vaccine composition comprising a virus-
CC like or a core particle with at least one first attachment site and at
CC least one antigen or antigenic determinant that is a prion protein (PrP)
CC or its dimer, or a PrP peptide, the antigen or antigenic determinant
CC being bound to the virus-like or core particle. The vaccine of the
CC invention may have neuroprotective or antiinflammatory activity. The
CC composition is useful as a medicament or in manufacturing a medicament
CC for the treatment or prevention of prion diseases. The prion diseases may
CC include Bovine Spongiform Encephalopathy (BSE) or Creutzfeldt-Jakob
CC Disease. The present sequence is that of a linker peptide which is
CC related to the invention.

XX SQ Sequence 18 AA;

Query Match 100.0%; Score 56; DB 7; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10

DB 1 PKPSTPPGSS 10

RESULT 14

AAY91034

ID AAY91034 standard; peptide; 42 AA.

XX AAY91034;

XX DT 06-SEP-2000 (first entry)

XX DE Lipid-tagged-scfv protein N-terminus peptide sequence.

XX KW Lipid tagged; LT-scfv; lipid modified proteinaceous molecule;
KW cell therapy; pharmaceutical; cellular membrane.

XX OS Unidentified.

XX PN WO200023570-A1.

XX PD 27-APR-2000.

XX PF 18-OCT-1999; 99WO-NL000644.

XX PR 16-OCT-1998; 98EP-00203482.

XX PA (UBIS-) U-BISYS BV.

XX PI Logtenberg T, De Kruij CA;

XX DR WPI; 2000-339673/29.

XX PT Altering the protein content of cellular membranes to produce
XX pharmaceutically active agents.

XX PS Disclosure; Page 20; 63pp; English.

XX CC The present invention describes a process (I) for modifying the protein
XX content of cellular membranes using lipid modified proteinaceous
XX molecules (lmpm). Cells and particles produced via (I) are used as
XX pharmaceuticals. For example they may be used for a cell therapy
XX protocol. (I) provides a novel approach to altering the biochemical
XX properties of cells (especially their ability to target tissues and
XX organs). It is a very rapid and efficient process and requires only small
XX amounts of lmpms which when integrated into the cells are stable in vivo.
XX (I) does not involve gene transfer (the protein is supplied directly to
XX the cells) and does not require the cells to be cultured after
XX integration of the protein. (I) may be applied to a wide range of cell
XX types not just primary human cells. The present sequence represents the N
XX -terminus peptide of lipid-tagged-scfv (LT-scfv) proteins expressed in
XX the vector pLP2, which is used in the exemplification of the present
XX invention

XX SQ Sequence 42 AA;

Query Match 100.0%; Score 56; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSTPPGSS 10

DB 30 PKPSTPPGSS 39

RESULT 15

AAR39337

ID AAR39337 standard; protein; 46 AA.

XX AC AAR39337;

XX DT 25-MAR-2003 (revised)

XX DT 01-FEB-1994 (first entry)

XX DE Intercalating GCN4-leucine zipper.

XX KW Expression; single-chain Fv; scFv; pLISC-SE; leucine zipper; linker;
KW hinge; IgG3; GCN4; cassette; restriction site; intercalating peptide;
KW intercalation.

OS Synthetic.

XX Key Location/Qualifiers
 XX FH 4..13
 FT Region /label= IGH3-hinge
 FT Protein 14..46
 FT FT /label= GCN4-zipper

XX WO9315210-A1.

XX PN

XX XX

XX PD 05-AUG-1993.

XX XX

XX PF 15-JAN-1993; 93WO-EP000082.

XX XX

XX PR 23-JAN-1992; 92EP-00101069.

XX XX

XX PA (MERE) MERCK PATENT GMBH.

XX XX

XX PI Plueckthun A, Pack P;

XX XX

XX DR WPI; 1993-258685/32.

XX DR N-PSDB; AAQ46824.

XX XX

XX FT Monomeric and dimeric antibody fragment fusion proteins - that use Fv

XX PT fragments of antibody but not constant antibody domains.

XX XX

XX PS Example 2; Page 30; 44pp; English.

XX XX

XX CC Example 2 describes the construction of a gene cassette encoding

XX CC intercalating peptides of a leucine zipper. The gene cassette, fitted

XX CC with restriction sites to be compatible with the restriction site at the

XX CC 3' end of the single-chain (sc) Fv fragment gene, must encode the

XX CC sequence of a hinge (connecting the scFv fragment to the intercalating

XX CC peptide) and the intercalating peptide itself. The hinge region, may

XX CC however be omitted. As an example, the sequence of the upper hinge region

XX CC of mouse IgG3, followed by the sequence of the leucine zipper sequence of

XX CC yeast protein GCN4 is back-translated into frequently used E. coli codons

XX CC (AAQ46824). Oligonucleotides are synthesised and ligated into pLISC-SB

XX CC (AAQ46823), previously digested with EcoRI and HindIII. (Updated on 25-

XX CC MAR-2003 to correct PN field.)

XX XX

XX SQ Sequence 46 AA;

Query Match

Best Local Similarity 100.0%; Score 56; DB 2; Length 46;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKPSTPPGSS 10

Db 4 PKPSTPPGSS 13

Search completed: April 20, 2004, 10:25:03

Job time : 41.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:39 ; Search time 44.4 Seconds
(without alignments)
106.594 Million cell updates/sec

Title: US-08-930-480A-5

Perfect score: 84

Sequence: 1 GGGGGGGGGGGGGGS 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	100.0	170	11 Q925S2	Q925B2 mus musculus
2	84	100.0	241	11 Q921A6	Q921A6 mus musculus
3	84	100.0	243	11 Q7TQM2	Q7TQM2 mus musculus
4	84	100.0	298	11 Q9QYF0	Q9QYF0 mus musculus
5	84	100.0	738	5 O02402	O02402 pinctada fu
6	80	95.2	592	16 Q9PF60	Q9PF60 xyliella fas
7	78	92.9	218	11 Q925S1	Q925S1 mus musculus
8	77	91.7	155	5 Q9VZK6	Q9VZK6 drosophila
9	76	90.5	80	10 Q9SUF7	Q9SUF7 arabidopsis
10	76	90.5	100	5 Q8MU90	Q8MU90 oncopeltus
11	76	90.5	104	5 Q9GN84	Q9GN84 drosophila
12	76	90.5	104	5 Q9GN83	Q9GN83 drosophila
13	76	90.5	113	10 Q8VY68	Q8VY68 arabidopsis
14	76	90.5	118	5 Q9VYS6	Q9VYS6 arabidopsis
15	76	90.5	158	5 Q9VYD8	Q9VYD8 drosophila
16	76	90.5	175	10 Q9LSN6	Q9LSN6 arabidopsis

17	76	90.5	204	5 Q9W2A7	Q9W2A7 drosophila
18	76	90.5	207	10 Q43522	Q43522 lycopersico
19	76	90.5	221	10 Q65514	Q65514 arabidopsis
20	76	90.5	222	10 Q7XDV2	Q7XDV2 oryza sativ
21	76	90.5	227	10 Q84W21	Q84W21 arabidopsis
22	76	90.5	255	10 Q9SIH2	Q9SIH2 arabidopsis
23	76	90.5	258	16 Q8XLQ8	Q8XLQ8 clostridium
24	76	90.5	259	5 Q9V6A4	Q9V6A4 drosophila
25	76	90.5	264	13 Q9DFB6	Q9DFB6 gallus gall
26	76	90.5	272	16 Q7W2S7	Q7W2S7 bordetella
27	76	90.5	280	10 Q7XU62	Q7XU62 oryza sativ
28	76	90.5	283	13 Q8AVB5	Q8AVB5 xenopus lae
29	76	90.5	284	4 Q9H524	Q9H524 homo sapien
30	76	90.5	288	10 Q7YI20	Q7YI20 oryza sativ
31	76	90.5	296	10 Q8RUS0	Q8RUS0 arabidopsis
32	76	90.5	300	16 Q7W0G7	Q7W0G7 bordetella
33	76	90.5	304	16 Q7WDS5	Q7WDS5 bordetella
34	76	90.5	321	6 Q9MYX6	Q9MYX6 ovnis aries
35	76	90.5	333	11 Q8C8L2	Q8C8L2 mus musculus
36	76	90.5	333	11 Q7TNS5	Q7TNS5 mus musculus
37	76	90.5	335	10 Q65330	Q65330 elaeagnus u
38	76	90.5	418	5 Q9W2R6	Q9W2R6 drosophila
39	76	90.5	447	13 Q73628	Q73628 anolis caro
40	76	90.5	493	11 Q8K1I7	Q8K1I7 mus musculus
41	76	90.5	528	4 Q13344	Q13344 homo sapien
42	76	90.5	541	16 Q87BZ7	Q87BZ7 xyliella fas
43	76	90.5	556	5 Q9VND4	Q9VND4 drosophila
44	76	90.5	646	5 Q9VI21	Q9VI21 drosophila
45	76	90.5	647	12 Q65013	Q65013 aleutian mi

ALIGNMENTS

RESULT 1

Q925S2 PRELIMINARY; PRT; 170 AA.
ID Q925S2
AC Q925S2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE MRP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAIB/c;
RX PubMed=11819679;
RA Cui D., Zeng G., Yan X., Wang F., Tian F., Ren D., Zhao T., Li X., Su C.;
RA "Mechanism of exogenous nucleic acids and their precursors improving the repair of intestinal epithelium after irradiation in mice.";
RL World J. Gastroenterol. 6:709-717(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BAIB/c;
RA Cui D., Zeng G., Yan X., Li X., Su C.;
RA "Cloning of mouse genes related to repairing of intestinal epithelium of the irradiated mice by treatment with the intestinal RNA of mice of the same strain.";
RL Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 19:71-80(2001).
DR EMBL; AF240167; AAK43732.1; -;
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
SQ SEQUENCE 170 AA; 17978 MW; 5042823CGC10F38 CRC64;
Query Match 100.0%; Score 84; DB 11; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 124 GGGGGGGGGGGGGG 138
 |||||

RESULT 2

Q921A6 PRELIMINARY; PRT; 241 AA.
 AC Q921A6; (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Anti-CEA 79 single chain Fv fragment (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98170165; PubMed=9509426;
 RA Chung J.H., Choi S.J., Kim H.J., Kim I.J., Choi I.H., Lee S.D.,
 RA Yi K.S., Suh P.G., Ryu S.H., Chung H.K.;
 RT "Cloning and characterization of cDNAs encoding VH and VL of a
 RT monoclonal anti-CEA antibody (CEA 79) cross-reactive with NCA-95 and
 RT generation of a single-chain Fv molecule (scFv).";
 RL Mol. Cells 7:816-819 (1997).
 DR EMBL: U88067; ABA8044.1; -.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig; 2.
 DR SMART: SM00406; IGV; 2.
 DR PROSITE: PS50835; IG_LIKE; 2.
 FT NON_TER 1 241
 FT NON_TER 241 241
 SQ SEQUENCE 241 AA; 25086 MW; 02768872489C771 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 241;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 119 GGGGGGGGGGGGGG 133
 |||||

RESULT 3

Q7TQM2 PRELIMINARY; PRT; 243 AA.
 AC Q7TQM2;
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE scFv 6H8 protein (Fragment).
 GN SCFV 6H8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Balb/c;
 RA Peter J.C., Eftekhari P., Billiald P., Wallukat G.;
 RT "scFv single chain antibody variable fragment as inverse agonist for
 RT the beta-2 adrenergic receptor".
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ574851; CAB00495.1; -.
 FT NON_TER 1 1
 FT NON_TER 1 1
 SQ SEQUENCE 243 AA; 25976 MW; BEFFF64D2DCF4F76 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 243;
 Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 117 GGGGGGGGGGGGGG 131
 |||||

RESULT 4

Q9QYF0 PRELIMINARY; PRT; 298 AA.
 AC Q9QYF0;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE CN 8 scFv.
 GN CN 8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Balb/c; TISSUE=Spleen;
 RX MEDLINE=20183931; PubMed=10706631;
 RA Shinohara N., Demura T., Fukuda H.;
 RT "Isolation of a vascular cell wall-specific monoclonal antibody
 RT recognizing a cell polarity by using a phage display subtraction
 RT method.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2585-2590 (2000).
 DR EMBL: AB036341; BAA88633.1; -.
 DR PIR: A33933; A33933.
 DR PIR: S19112; S19112.
 DR HSSP: P01607; IREI.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig; 2.
 DR SMART: SM00406; IGV; 2.
 DR PROSITE: PS50835; IG_LIKE; 2.
 SQ SEQUENCE 298 AA; 31867 MW; E0F96BBA17004317 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 298;
 Best Local Similarity 100.0%; Pred. No. 0.052;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 158 GGGGGGGGGGGGGG 172
 |||||

RESULT 5

Q02402 PRELIMINARY; PRT; 738 AA.
 AC Q02402;
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Insoluble protein.
 OS Pinctada fucata.
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;
 OC Pterioidea; Pteriidae; Pinctada.
 OX NCBI_TaxID=50426;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97320490; PubMed=9177341;
 RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
 RA Nakashima K., Takahashi T.;
 RT "Structures of mollusc shell framework proteins.";
 RL Nature 387:563-564 (1997).
 DR EMBL: D86074; BAA20466.1; -.
 SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 100.0%; Score 84; DB 5; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGG 15
 Db 666 GGGGGGGGGGGG 680

RESULT 6
 Q9PF60 PRELIMINARY; PRT; 592 AA.

ID Q9PF60
 AC Q9PF60
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Endo-1,4-beta-glucanase.
 GN XP0818.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemp E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E., Laigret F., Lamais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Sawasaki H.E.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL; AF003921; AAF83628.1; --
 DR PIR; E82759; E82759.
 DR HSP; P54583; IECE.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001919; Bac celose-bind.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00553; CBM 2; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR SMART; SM00637; CBD II; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Complete proteome.
 SQ SEQUENCE 592 AA; 59967 MW; 9846D4EA3B5C89E CRC64;

Query Match 95.2%; Score 80; DB 16; Length 592;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGG 14
 Db 1 GGGGGGGGGGGG 14

Db 467 GGGGGGGGGGGG 480

RESULT 7
 Q925S1 PRELIMINARY; PRT; 218 AA.

ID Q925S1
 AC Q925S1
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MRP5 (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RX PubMed=11819679;
 RA Cui D., Zeng G., Yan X., Wang F., Tian F., Ren D., Zhao T., Li X.,
 RA Su C.;
 RT "Mechanism of exogenous nucleic acids and their precursors improving
 the repair of intestinal epithelium after irradiation in mice";
 RL World J. Gastroenterol. 6:709-717(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Cui D., Zeng G., Yan X., Li X., Su C.;
 RT "Cloning of mouse genes related to repairing of intestinal epithelium
 of the irradiated mice by treatment with the intestinal RNA of mice of
 the same strain";
 RL Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 19:71-80(2001).
 DR EMBL; AF240168; AAK43733.1; --
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig; 2.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 FT NON TER 218
 SQ SEQUENCE 218 AA; 23013 MW; 527E4FA8F7982817 CRC64;

Query Match 92.9%; Score 78; DB 11; Length 218;
 Best Local Similarity 93.3%; Pred. No. 0.17;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGG 15
 Db 121 GGGGGGGGGGGG 135

RESULT 8
 Q9VZK6 PRELIMINARY; PRT; 155 AA.

ID Q9VZK6
 AC Q9VZK6
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE CG10853 protein (LP09837p).
 GN CG10853 OR BCDNA:LP09837.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolenakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiambo I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Calniker S.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003479; AAF47815.1; -;
 DR EMBL: AY075437; AAL68252.1; -;
 DR FLYBase: FBgn0035478; CGI0853.
 SQ SEQUENCE 155 AA; 14855 MW; EF7D78EDD16675BF CRC64;

Query Match 91.7%; Score 77; DB 5; Length 155;
 Best Local Similarity 86.7%; Pred. No. 0.15;
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGGGS 15
 Db 58 GGGGGGGGGGGGGGA 72

RESULT 9

Q9SUF7 PRELIMINARY; PRT; 80 AA.
 AC Q9SUF7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE Hypothetical protein.
 GN T12G13.70 OR A74G08230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,
 RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,
 RA Schueller C.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL080252; CAB45793.1; -;
 DR EMBL: AL161510; CAB81159.1; -;
 DR PIR: T10550; T10550.
 KW Hypothetical protein.
 SQ SEQUENCE 80 AA; 7872 MW; A1BEE43FCA7ED68 CRC64;

Query Match 90.5%; Score 76; DB 10; Length 80;
 Best Local Similarity 92.9%; Pred. No. 0.1;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
 Db 22 GGGGGGGGGGGGGG 35

RESULT 10

Q8MU90 PRELIMINARY; PRT; 100 AA.
 ID Q8MU90;
 AC Q8MU90;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Tiptop transcription factor (Fragment).
 GN TIPTOP.
 OS Oncopeltus fasciatus (Milkweed bug).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Heteroptera;
 OC Panheteroptera; Pentatomomorpha; Lygaeoidea; Lygaeidae; Lygaeinae;
 OC Oncopeltus.
 OX NCBI_TaxID=7536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rogers B.T., Herke S.W.;
 RT "partial tiptop cDNA from *Oncopeltus fasciatus*";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF533539; AAM97356.1; -;
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 1.
 DR SMART: SM00355; Znf_C2H2; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 1.
 KW Metal-binding; Zinc; Zinc-finger.
 FT NON_TER 1
 FT NON_TER 100
 SQ SEQUENCE 100 AA; 9919 MW; 650D5E401FEF35CD CRC64;

Query Match 90.5%; Score 76; DB 5; Length 100;
 Best Local Similarity 86.7%; Pred. No. 0.13;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGGGS 15
 Db 65 GGGGGGGGGGGGGGS 79

RESULT 11

```

Q9GN84
ID Q9GN84 PRELIMINARY; PRT; 104 AA.
AC Q9GN84;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NONA protein (Fragment).
GN NONA.
OS Drosophila littoralis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]_TaxID=47316;
RP SEQUENCE FROM N.A.
RC STRAIN=Ru3, and Ru1;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304367; CAC20090.1; --
DR EMBL; AJ304361; CAC20084.1; --
DR FlyBase; FBgn0043410; Dlit\nona.
FT NON_TER 104
SQ SEQUENCE 104 AA; 10048 MW; AC804E039196298C CRC64;

Query Match 90.5%; Score 76; DB 5; Length 104;
Best Local Similarity 92.9%; Pred. No. 0.13;
Matches 13; Conservative 0; Mismatches 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
Db 68 GGGGGGGGGGGGGG 81

RESULT 12
Q9GN83
ID Q9GN83 PRELIMINARY; PRT; 104 AA.
AC Q9GN83;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NONA protein (Fragment).
GN NONA.
OS Drosophila littoralis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]_TaxID=47316;
RP SEQUENCE FROM N.A.
RC STRAIN=Sa5, Ou5, Ku5, Sal, and Sa3;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304352; CAC20075.1; --
DR EMBL; AJ304316; CAC20039.1; --
DR EMBL; AJ304334; CAC20057.1; --
DR EMBL; AJ304340; CAC20063.1; --
DR EMBL; AJ304346; CAC20069.1; --
DR FlyBase; FBgn0043410; Dlit\nona.
FT NON_TER 104
SQ SEQUENCE 104 AA; 10047 MW; AC9EAB039196298C CRC64;

Query Match 90.5%; Score 76; DB 5; Length 104;
Best Local Similarity 92.9%; Pred. No. 0.13;
Matches 13; Conservative 0; Mismatches 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
Db 68 GGGGGGGGGGGGGG 81

Q9GN84
ID Q9GN84 PRELIMINARY; PRT; 113 AA.
AC Q9GN84;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN AF408230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]_TaxID=3702;
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayaehizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayaehizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072405; RAL62398.1; --
DR EMBL; BT000222; AAN15541.1; --
KW Hypothetical protein.
SQ SEQUENCE 113 AA; 11548 MW; D0182159545EF3F9 CRC64;

Query Match 90.5%; Score 76; DB 10; Length 113;
Best Local Similarity 92.9%; Pred. No. 0.14;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
Db 63 GGGGGGGGGGGGGG 76

RESULT 14
Q9VYS6
ID Q9VYS6 PRELIMINARY; PRT; 118 AA.
AC Q9VYS6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG1840 protein (LD12750P).
GN CG1840 OR BCDNA:LD12750.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_TaxID=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Perrier S., Fleischmann W.,
RA Foeiler K.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarri C., McLeod M.P., McPherson D.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003487; AAF48112.1; -;
DR EMBL; AY094783; AAM11136.1; -;
DR FlyBase; FBgn0030351; CG1840.
SQ SEQUENCE 118 AA; 11803 MW; 0568ACA6501716AC CRC64;
Query Match 90.5%; Score 76; DB 5; Length 118;
Best Local Similarity 86.7%; Pred. No. 0.15;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGGGGGGGGGGGGGS 15
Db 74 GGGGGGGGGGGGGGS 88
RESULT 15
QYVD8 PRELIMINARY; PRT; 158 AA.
AC QYVD8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG1987 protein (RE47308P).
GN RBP1-LIKE OR CG1987.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

ANANATIDES P.G., SCHERER S.E., LI P.W., HOSKINS R.A., GALLE R.F.,
GEORGE R.A., LEWIS S.E., RICHARDS S., ASHBURNER M., HENDERSON S.N.,
SUTTON G.G., WORTMAN J.R., YANDELL M.D., ZHANG Q., CHEN L.X.,
BRANDON R.C., ROGERS Y.-H.C., BLAZER R.G., CHAMPE M., PFEIFFER B.D.,
WAN K.H., DOYLE C., BAXTER E.G., HEIT G., NELSON C.R., MIKLOS G.L.G.,
ARIL J.F., AGBAYANI A., AN H.-J., ANDREWS-FRANKOCH C., BALDWIN D.,
BAILEY R.M., BASU A., BAXENDALE J., BAYRAKTAROGU L., BEASLEY E.M.,
BEESON K.Y., BENOS P.V., BERMAN B.P., BHANDARI D., BOLSHAKOV S.,
BORKOVA D., BOTCHAN M.R., BOUCK J., BROKSTEIN P., BROTTIER P.,
BURTIS K.C., BUSAM D.A., BUTLER H., CADIEU E., CENTER A., CHANDRA I.,
CHERRY J.M., CAWLEY S., DAHLKE C., DAVENPORT L.B., DAVIES P.,
DE PABLOS B., DELCHER A., DENG Z., MAYS A.D., DEW I., DIETZ S.M.,
DODSON K., DOUP L.E., DOWNES M., DUGAN-ROCHA S., DUNKOV B.C., DUNN P.,
DURBIN K.J., EVANGELISTA C.C., FERRAZ C., FERRIERA S., FLEISCHMANN W.,
FOELER K.J., GABRIELIAN A.E., GARG N.S., GELBART W.M., GLASSER K.,
GLODEK A., GONG F., GORRELL J.H., GU Z., GUAN P., HARRIS M.,
HARRIS N.L., HARVEY D., HEIMAN T.J., HERNANDEZ J.R., HOUCK J.,
HOSTIN D., HOUSTON K.A., HOWLAND T.J., HERNANDEZ J.R., HOUCK J.,
JALALI M., KALUSH F., KARPEN G.H., KE Z., KENNISON J.A., KETCHUM K.A.,
KIMMEL B.E., KODIRA C.D., KRAFT C., KRAVITZ S., KULP D., LAI Z.,
LASKO P., LEI Y., LEVITSKY A.A., LI J., LI Z., LIANG Y., LIN X.,
LIU X., MATTEI B., MCINTOSH T.C., MCLEOD M.P., MCPHERSON D.,
MERKULOV G., MILSHINA N.V., MOBARRY C., MORRIS J., MOSHREFI A.,
MOUNT S.M., MOY M., MURPHY B., MURPHY L., MUZNY D.M., NELSON D.L.,
NELSON D.R., NELSON K.A., NIXON K., NUSSKERN D.R., PACLEB J.M.,
PALAZZOLO M., PITTMAN G.S., PAN S., POLLARD J., PURI V., REESE M.G.,
REINERT K., REMINGTON K., SAUNDERS R.D.C., SCHEELER F., SHEN H.,
SHUE B.C., SIDEN-KIAMOS I., SIMPSON M., SKUPSKI M.P., SMITH T.,
SPIER E., SPRADLING A.C., STAPLETON M., STRONG R., SUN E.,
SWIRSKAS R., TECTOR C., TURNER R., VENTER E., WANG A.H., WANG X.,
WANG Z.-Y., WASSARMAN D.A., WEINSTEIN G.M., WEISSENBACK J.,
WILLIAMS S.M., WOODAGE T., WORLEY K.C., WU D., YANG S., YAO Q.A.,
YE J., YEH R.-F., ZAVERI J.S., ZHAN M., ZHANG G., ZHAO Q., ZHENG L.,
ZHENG X.H., ZHONG F.N., ZHONG W., ZHOU X., ZHU S., ZHU X., SMITH H.O.,
GIBBS R.A., MYERS E.W., RUBIN G.M., VENTER J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananattides P.G., Brandon R.C., Rogers Y.,
RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleeb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003492; AAF48264.2; -;
 DR EMBL; AY113490; AAM29495.1; -;
 DR FlyBase; FBgn0030479; Rbpl-like.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 SQ SEQUENCE 158 AA; 16799 MW; C6D065ABD640EAE5 CRC64;

Query Match 90.5%; Score 76; DB 5; Length 158;
 Best Local Similarity 92.9%; Pred. No. 0.2;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
 Db 94 GGGGGGGGGGGGGG 107

Search completed: April 20, 2004, 10:27:02
 Job time : 46.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:04 ; Search time 12.6 Seconds
(without alignments)

61.988 Million cell updates/sec

Title: US-08-930-480A-5

Perfect score: 84

Sequence: 1 GGGGGGGGGGGGGGS 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	80	95.2	306	1 RALY_HUMAN	Q9ukm9 homo sapien
2	79	94.0	104	1 HOL3_HOLDI	Q25055 holotrichia
3	77	91.7	206	1 TWS1_MOUSE	P26887 mus musculus
4	76	90.5	165	1 GRP1_ORYSA	P25074 oryza sativ
5	76	90.5	280	1 CHIA_MAIZE	P29022 zea mays (m
6	76	90.5	321	1 PUR_MOUSE	P42669 mus musculus
7	76	90.5	322	1 PUR_HUMAN	Q00577 homo sapien
8	76	90.5	378	1 RU17_MOUSE	P09026 mus musculus
9	76	90.5	433	1 HXB3_MOUSE	Q24563 drosophila
10	76	90.5	539	1 DOP2_DROME	Q9hcs4 homo sapien
11	76	90.5	588	1 T7LI_HUMAN	Q00004 canis famil
12	76	90.5	622	1 SR68_CANFA	Q09112 mus musculus
13	76	90.5	663	1 DUS8_MOUSE	O54839 mus musculus
14	76	90.5	688	1 EOMD_MOUSE	P22670 homo sapien
15	76	90.5	979	1 RFX1_HUMAN	O42131 gallus gall
16	76	90.5	1627	1 TP2B_CHICK	Q15911 homo sapien
17	76	90.5	3703	1 ABF1_HUMAN	Q61329 mus musculus
18	76	90.5	3726	1 ABF1_MOUSE	P06813 oryctolagus
19	74	88.1	266	1 CANS_RABIT	P70390 mus musculus
20	74	88.1	331	1 SHX2_MOUSE	Q98937 gallus gall
21	74	88.1	440	1 FXG3_CHICK	Q61060 mus musculus
22	74	88.1	465	1 FXD3_MOUSE	Q9ub98 homo sapien
23	74	88.1	627	1 SR68_HUMAN	P03136 hamster par
24	74	88.1	722	1 COAT_PAVHH	Q9Y2x9 homo sapien
25	74	88.1	895	1 Z2B1_HUMAN	P43029 mus musculus
26	73	86.9	151	1 GDF7_MOUSE	P13135 bos taurus
27	73	86.9	263	1 CANS_BOVIN	P04574 sus scrofa
28	73	86.9	266	1 CANS_PIG	P04632 homo sapien
29	73	86.9	268	1 CANS_HUMAN	O09029 mesocricetu
30	73	86.9	367	1 BET3_MESAU	Q9td03 arabidopsis
31	73	86.9	377	1 HSF7_ARATH	O00570 homo sapien
32	73	86.9	387	1 SOX1_HUMAN	P53783 mus musculus
33	73	86.9	391	1 SOX1_MOUSE	

RESULT 1
RALY_HUMAN
ID RALY_HUMAN STANDARD; PRT; 306 AA.
AC Q9UKM9; Q14621; Q9EQX6; Q9UJE3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-binding protein Raly (hnRNP associated with lethal yellow homolog)
DE (Autoantigen p542).
GN RALY OR P452.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RX MEDLINE=99431566; PubMed=10500250;
RA Khrebtukova I., Kuklin A., Woychik R.P., Michaud E.J.;
RT "Alternative processing of the human and mouse raly genes.";
RL Biochim. Biophys. Acta 1447:107-112(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Vaughan J.H.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
MEDLINE=21638749; PubMed=11780052;
RA Jones M., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Deloukas P., Matthews G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Clegg S., Cobley V.E., Collier R.E., Clark L.N., Clark S.Y., Clee C.M.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.S.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromana A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [4]

O75444 homo sapien
Q03828 homo sapien
P31361 mus musculus
Q63262 rattus norv
P20264 homo sapien
Q81X10 homo sapien
P19289 thermoprote
P26968 tenebrio mo
Q13595 homo sapien
P29031 populus tri
O60902 homo sapien
O61374 ceratitis c

SEQUENCE OF 85-306 FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
RP TISSUE=Lymphocytes;
RX MEDLINE=98018738; PubMed=9376072;
RA Rhodes G.H., Valbracht J.R., Nguyen M.-D., Vaughan J.H.;
RT "The p542 gene encodes an autoantigen that cross-reacts with EBNA-1 of
the Epstein Barr virus and which may be a heterogeneous nuclear
ribonucleoprotein."
RL J. Autoimmun. 10:447-454 (1997).
RN [5]
RP SEQUENCE OF 227-253, AND DETERMINATION OF AUTOANTIGENIC EPITOPE.
RX MEDLINE=95190029; PubMed=7533788;
RA Vaughan J.H., Valbracht J.R., Nguyen M.-D., Handley H.H., Smith R.S.,
Patrick K., Rhodes G.H.;
RT "Epstein-Barr virus-induced autoimmune responses. I. Immunoglobulin M
autoantibodies to proteins mimicking and not mimicking Epstein-Barr
virus nuclear antigen-1."
RL J. Clin. Invest. 95:1306-1315 (1995).
CC -!- FUNCTION: Probable-RNA binding protein. Could be a heterogeneous
nuclear ribonucleoprotein (hnRNP).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2;
CC IsoId=Q9UKM9-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9UKM9-2; Sequence=VSP_005804;
CC -!- TISSUE SPECIFICITY: Expressed in heart, brain, lung, liver,
skeletal muscle, kidney and pancreas. Weakly expressed in
placenta.
CC -!- DISEASE: Autoantigen found in infectious mononucleosis caused by
Epstein-Barr virus. An epitope recognized by B-cells, which cross-
react with the BKRF1 protein (EBNA-1 nuclear protein) of Epstein-
Barr virus has been identified.
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -!- CAUTION: Ref.4 (CAC29371) sequence differs from that shown due to
erroneous gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF148457; AAC04487.1; -;
CC EMBL; L38696; AAC28898.1; -;
CC EMBL; AL031668; CAC29371.1; ALT_SEQ.
CC EMBL; AL031668; CAB43742.1; -;
CC Genew; HGNC:15921; RALY.
CC InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 1.
CC SMART; SM00360; RRM; 1.
CC DR PROSITE; PS0102; RRM; 1.
CC DR PROSITE; PS00030; RRM_RNP_1; 1.
CC KW Ribonucleoprotein; RNA-binding; Nuclear protein; Antigen;
KW Alternative splicing; Polymorphism.
FT DOMAIN 21 92 RNA-BINDING (RRM).
FT 227 253 EPITOPE (RECOGNIZED BY BKRF1 ANTIBODIES).
FT DOMAIN 225 251 POLY-GLY.
FT VARSPLIC 110 125 Missing (in isoform 1).
FT VARIANT 215 215 Q -> R (in dbSNP:3180568).
FT FTId=VAR_015223.
FT VARIANT 251 251 G -> S (in dbSNP:2281209).
FT FTId=VAR_015224.
FT CONFLICT 214 215 EQ -> DE (IN REF. 2).
FT CONFLICT 230 230 A -> AS (IN REF. 2).
SQ SEQUENCE 306 AA; 32463 MW; 7F4376D3BD8E4728 CRC64;
Query Match 95.2%; Score 80; DB 1; Length 306;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
Db 235 GGGGGGGGGGGGGG 248
RESULT 2
HOL3 HOLDI STANDARD; PRT; 104 AA.
ID AC Q25055;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holotricin 3 precursor.
OS Holotrichia diomphalia.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia;
OC Scarabaeidae; Melolonthinae; Holotrichia.
OX NCBI_TaxID=33394;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-40.
RX TISSUE=Larval hemolymph;
RX MEDLINE=96073722; PubMed=8535393;
RA Lee S.Y., Moon H.J., Kurata S., Natori S., Lee B.L.;
RT "Purification and cDNA cloning of an antifungal protein from the
hemolymph of Holotrichia diomphalia larvae."
RL Biol. Pharm. Bull. 18:1049-1052 (1995).
CC -!- FUNCTION: Has antifungal activity against C.albicans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: TO TENECIN 3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D13744; BAA02889.1; -;
CC PIR; JC4190; JC4190.
CC Insect immunity; Antibiotic; Hemolymph; Fungicide; Signal; Repeat.
FT SIGNAL 1 20
FT CHAIN 21 104
FT DOMAIN 27 98
FT REPEAT 27 30 1.
FT REPEAT 31 34 2.
FT REPEAT 35 38 3.
FT REPEAT 39 42 4.
FT REPEAT 43 46 5.
FT REPEAT 47 50 6.
FT REPEAT 51 54 7.
FT REPEAT 55 58 8.
FT REPEAT 59 62 9.
FT REPEAT 63 66 10.
FT REPEAT 67 70 11.
FT REPEAT 71 74 12.
FT REPEAT 75 78 13.
FT REPEAT 79 82 14.
FT REPEAT 83 86 15.
FT REPEAT 87 90 16.
FT REPEAT 91 94 17.
FT REPEAT 96 98 18.
SQ SEQUENCE 104 AA; 9026 MW; 2799D681BFDCC725 CRC64;
Query Match 94.0%; Score 79; DB 1; Length 104;
Best Local Similarity 93.3%; Pred. No. 0.061;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGGGGGGGGGGGG 15
Db 64 GGGGGGGGGGGGGG 78

```

CC CC bHLH protein. Homodimer.
CC CC -!- SUBCELLULAR LOCATION: Nuclear.
CC CC -!- TISSUE SPECIFICITY: Subset of mesodermal cells.
CC CC -!- INDUCTION: By TGF-alpha
CC CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch)
CC CC -----
CC DR EMBL; MG3649; AAA40514.1; -
CC DR EMBL; MG3650; AAA40515.1; -
CC DR EMBL; BC033434; AAH33434.1; -
CC DR PIR; I53066; I53066.
CC DR TRANSFAC; T01635; -
CC DR MGD; MGI:98872; Twist1.
CC DR GO; GO:0005634; C:nucleus; NAS.
CC DR GO; GO:0003700; F:transcription factor activity; NAS.
CC DR GO; GO:0030154; P:cell differentiation; IMP.
CC DR GO; GO:0030326; P:limb morphogenesis; IMP.
CC DR GO; GO:0045596; P:negative regulation of cell differentiation; IDA.
CC DR GO; GO:0045843; P:negative regulation of myogenesis; IDA.
CC DR GO; GO:0001679; P:neurulation; IMP.
CC DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
CC DR InterPro; IPR001092; HLH_basic.
CC DR Pfam; PF00010; HLH; 1.
CC DR SMART; SM00353; HLH; 1.
CC DR PROSITE; PS05888; HLH; 1.
CC DR Transcription; Developmental protein; Nuclear protein; DNA-binding;
CC KW Differentiation; Developmental protein; Nuclear protein; DNA-binding;
CC FT DOMAIN 90 102 GLY-RICH.
CC FT DNA_BIND 112 124 BASIC DOMAIN.
CC FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF.
CC FT VARIANT 36 36 A -> R (IN CDNA).
CC FT VARIANT 91 91 G -> P (IN CDNA).
CC SQ SEQUENCE 206 AA; 21198 MW; 618AD9E9B87C555 CRC64;
CC Query Match 91.7%; Score 77; DB 1; Length 206;
CC Best Local Similarity 86.7%; Pred. No. 0.17;
CC Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC OY 1 GGGGGGGGGGGGGGGG 15
CC DB 83 GGGGGGGGGGGGGGGG 97
CC -----
CC RESULT 4
CC ID GRP1 ORYSA STANDARD; PRT; 165 AA.
CC AC P25074;
CC DT 01-MAY-1992 (Rel. 22, Created)
CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
CC DT 01-APR-1993 (Rel. 25, Last annotation update)
CC DE Glycine-rich cell wall structural protein 1 precursor.
CC GN GRP-1.
CC OS Oryza sativa (Rice).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
CC OC Ehrhartoideae; Oryzaceae; Oryza.
CC OX NCBI_TaxID=4530;
CC RN [1]_TaxID=4530;
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=cv. Indica-IR36;
CC RX MEDLINE=91370862; PubMed=1716496;
CC RA Lei M., Wu R.;
CC RT "A novel glycine-rich cell wall protein gene in rice.";
CC RL Plant Mol Biol. 16:187-198(1991).
CC -!- FUNCTION: Responsible for plasticity of the cell wall (Potential).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X53596; CAA37665.1; --
 DR PIR; S13385; KNRZG1.
 DR Gramene; F25074; --
 KW Cell wall; Structural protein; Repeat; Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL
 FT PROTEIN 1.
 FT DOMAIN 31 159 GLY-RICH.
 FT REPEAT 56 62 R2 (TYR-RICH).
 FT REPEAT 93 99 R2 (TYR-RICH).
 FT REPEAT 132 138 R2 (TYR-RICH).
 SQ SEQUENCE 165 AA; E36CE31C3650AC9A CRC64;
 Query Match 90.5%; Score 76; DB 1; Length 165;
 Best Local Similarity 86.7%; Pred. No. 0.18;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGGGGGGGGGGGGGS 15
 |||||
 Db 111 GGGGGGGGGGGGGGS 125
 RESULT 5
 CHIA-MAIZE
 ID CHIA-MAIZE STANDARD; PRT; 280 AA.
 AC P29022;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Endochitinase A precursor (EC 3.2.1.14) (Seed chitinase A).
 OS Zea mays (Maize)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OC NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed;
 RX MEDLINE=92202208; PubMed=1551872;
 RA Huynh Q.K., Hironaka C.M., Levine E.B., Smith C.E., Borgmeyer J.R.,
 RA Shah D.M.;
 RT "Antifungal proteins from plants. Purification, molecular cloning,
 RT and antifungal properties of chitinases from maize seed.";
 RL J. Biol. Chem. 267:6635-6640(1992).
 RN [2]
 RP SEQUENCE OF 180-195.
 RC TISSUE=Seed;
 RX MEDLINE=92156129; PubMed=1740436;
 RA Verburg J.G., Smith C.E., Lisek C.A., Huynh Q.K.;
 RT "Identification of an essential tyrosine residue in the catalytic
 RT site of a chitinase isolated from Zea mays that is selectively
 RT modified during inactivation with
 RT 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide.";
 RL J. Biol. Chem. 267:3886-3893(1992).
 RN [1]
 RP FUNCTION: This protein functions as a defense against chitin
 CC containing fungal pathogens.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
 CC acetyl-D-glucosamine polymers of chitin.
 CC -!- MISCELLANEOUS: Maize chitinase B seems to be less active than
 CC chitinase A.
 CC -!- SIMILARITY: Belongs to chitinase class IA (family 19 of glycosyl
 CC hydrolases).
 CC -!- SIMILARITY: Contains 1 chitin-binding type-1 domain.
 CC -----

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M84164; AAA33444.1; --
 DR HSPF; P02877; IHEV.
 DR MaizeDB; 25130; --
 DR InterPro; IPR001002; Chitin binding 1.
 DR InterPro; IPR000726; Glyco_Hydro_19.
 DR Pfam; PF00187; Chitin_bind_1.
 DR Pfam; PF00182; Glyco_Hydro_19; 1.
 DR PRINTS; P00451; CHITINBINDING.
 DR ProDom; PD000609; Chitin binding 1; 1.
 DR ProDom; PD354900; Glyco_Hydro_19; 1.
 DR SMART; SM00270; ChtBD1; 1.
 DR PROSITE; PS00941; CHIT_BIND_I_2; 1.
 DR PROSITE; PS00026; CHIT_BIND_I_1; 1.
 DR PROSITE; PS00773; CHITINASE_19; 1.
 DR PROSITE; PS00774; CHITINASE_19_2; 1.
 KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;
 KW Multigene family.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 280 ENDOCHITINASE A.
 FT DOMAIN 26 60 CHITIN-BINDING TYPE-1.
 FT DOMAIN 61 77 HINGE REGION (POLY-GLY).
 FT DOMAIN 78 280 CATALYTIC.
 FT DISULFID 28 36 BY SIMILARITY.
 FT DISULFID 30 42 BY SIMILARITY.
 FT DISULFID 35 49 BY SIMILARITY.
 FT DISULFID 53 58 BY SIMILARITY.
 FT DISULFID 100 149 BY SIMILARITY.
 FT DISULFID 161 170 BY SIMILARITY.
 FT DISULFID 248 280 BY SIMILARITY.
 SQ SEQUENCE 280 AA; 29124 MW; 4FC5BB7D938C1CC1 CRC64;
 Query Match 90.5%; Score 76; DB 1; Length 280;
 Best Local Similarity 86.7%; Pred. No. 0.28;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGGGGGGGGGGGGGS 15
 |||||
 Db 61 GGGGGGGGGGGGGGS 75
 RESULT 6
 PUR MOUSE
 ID PUR MOUSE STANDARD; PRT; 321 AA.
 AC P42669;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Transcriptional activator protein PUR-alpha (Purine-rich single-
 DE stranded DNA-binding protein alpha).
 GN PURA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95047494; PubMed=7959008;
 RA Ma Z.-W., Bergemann A.D., Johnson E.M.;
 RT "Conservation in human and mouse Pur alpha of a motif common to
 RT several proteins involved in initiation of DNA replication.";
 RL Gene 149:311-314(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=97476282; PubMed=9334258;

```

RA Kelm R.J. Jr., Elder P.K., Strauch A.R., Getz M.J.;
RT "Sequence of cDNAs encoding components of vascular actin
RT single-stranded DNA-binding factor 2 establish identity to Puralpha
RT and Purbeta.";
RL J. Biol. Chem. 272:26727-26733(1997).
CC
CC -!- FUNCTION: THIS IS A PROBABLE TRANSCRIPTION ACTIVATOR THAT
CC SPECIFICALLY BINDS THE PURINE-RICH SINGLE STRAND OF THE PUR
CC ELEMENT LOCATED UPSTREAM OF THE C-MYC GENE. MAY PLAY A ROLE IN
CC THE INITIATION OF DNA REPLICATION AND IN RECOMBINATION.
CC
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U02098; AAA64630.1; -.
CC EMBL; AF017631; AAB71860.1; -.
CC TRANSFAC; T05167; -.
CC MGD; MGI:103079; Pura.
CC InterPro; IPR006628; PUR_DNA_RNA.
CC Pfam; PF04845; Pura; 1.
CC SMART; SM00712; PUR; 3.
CC Transcription regulation; Activator; DNA-binding; Nuclear protein.
CC DOMAIN 11 52 GLN/GLU-RICH; PART OF THE TRANSCRIPTIONAL
CC DOMAIN 292 321 ACTIVATION DOMAIN (POTENTIAL).
CC SEQUENCE 321 AA; 34884 MW; 0379DBD96D47DCEA CRC64;
CC
CC Query Match 90.5%; Score 76; DB 1; Length 321;
CC Best Local Similarity 92.9%; Pred. No. 0.32;
CC Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC Qy 1 GGGGGGGGGGGGGG 14
CC ||||| ||||| |||||
CC Db 34 GGGGGGGGGGGGGG 47
CC
CC RESULT 7
CC PUR_HUMAN
CC ID PUR_HUMAN STANDARD; PRT; 322 AA.
CC AC Q00577;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Transcriptional activator protein PUR-alpha (Purine-rich single-
CC stranded DNA-binding protein alpha).
CC GN PURA OR PUR1.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Fetal liver;
CC RX MEDLINE=93078769; PubMed=1448097;
CC RA Bergemann A.D., Ma Z.-W., Johnson E.M.;
CC RT "Sequence of cDNA comprising the human pur gene and sequence-specific
CC single-stranded-DNA-binding properties of the encoded protein.";
CC RL Mol. Cell. Biol. 12:5673-5682(1992).
CC
CC -!- FUNCTION: This is a probable transcription activator that
CC specifically binds the purine-rich single strand of the PUR
CC element located upstream of the MYC gene. May play a role in
CC the initiation of DNA replication and in recombination.
CC
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M96684; AAA60229.1; -.
CC EMBL; U02098; -. NOT_ANNOTATED_CDS.
CC Genew; HGNC:9701; PURA.
CC MIM; 600473; -.
CC GO; GO:0003705; F:RNA polymerase II transcription factor acti. . .; TAS.
CC GO; GO:0003697; F:single-stranded DNA binding; TAS.
CC GO; GO:0006270; P:DNA replication initiation; TAS.
CC GO; GO:0006270; P:DNA replication initiation; TAS.
CC InterPro; IPR006628; PUR_DNA_RNA.
CC Pfam; PF04845; Pura; 1.
CC SMART; SM00712; PUR; 3.
CC Transcription regulation; Activator; DNA-binding; Nuclear protein.
CC DOMAIN 11 53 GLN/GLU-RICH.
CC DOMAIN 293 322 ACTIVATION DOMAIN (POTENTIAL).
CC SEQUENCE 322 AA; 34911 MW; 797968504D01B356 CRC64;
CC
CC Query Match 90.5%; Score 76; DB 1; Length 322;
CC Best Local Similarity 92.9%; Pred. No. 0.32;
CC Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC Qy 1 GGGGGGGGGGGGGG 14
CC ||||| ||||| |||||
CC Db 34 GGGGGGGGGGGGGG 47
CC
CC RESULT 8
CC RUI7_MOUSE
CC ID RUI7_MOUSE STANDARD; PRT; 378 AA.
CC AC Q62376;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 43, Last annotation update)
CC DE U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (snRNP70)
CC DE (Fragment).
CC GN SNRP70.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN [1]
CC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
CC STRAIN=BALB/c;
CC RX MEDLINE=89276388; PubMed=2525092;
CC RA Hornig H., Fischer U., Costas M., Rauh A., Luehrmann R.;
CC RT "Analysis of genomic clones of the murine U1RNA-associated 70-kDa
CC protein reveals a high evolutionary conservation of the protein
CC between human and mouse.";
CC RL Eur. J. Biochem. 182:45-50(1989).
CC -!- FUNCTION: Mediates the splicing of pre-mRNA by binding to the loop
CC I region of U1-snRNA. The truncated isoform cannot bind U1-snRNA
CC (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62376-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62376-2; Sequence=VSP_005851, VSP_005852;
CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE C-
CC TERMINAL REGION (BY SIMILARITY).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; X15769; CAA33777.1; --
DR EMBL; X15770; CAA33777.1; JOINED.
DR EMBL; X15771; CAA33777.1; JOINED.
DR EMBL; X15772; CAA33777.1; JOINED.
DR EMBL; X15773; CAA33777.1; JOINED.
DR EMBL; X15774; CAA33777.1; JOINED.
DR EMBL; X15775; CAA33777.1; JOINED.
DR EMBL; X15776; CAA33777.1; JOINED.
DR PIR; S04336; S04336.
DR PIR; S04824; S04824.
DR HSP; P09651; IHA1.
DR MGD; MGI:98341; Snrp70.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; xrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00102; RRM; 1.
DR PROSITE; PS00030; Ribonucleoprotein; RNA-binding; phosphorylation;
KW Nuclear protein; Ribonucleoprotein; RNA-binding; phosphorylation;
KW Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 33 111 RNA-BINDING (RRM).
FT DOMAIN 161 240 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 241 256 POLY-GLY.
FT DOMAIN 286 333 ARG/ASP/GLU-RICH (MIXED CHARGE).
FT DOMAIN 334 339 POLY-GLY.
FT VARSPLIC 90 96 AYKHADG -> TTQLACS (in isoform 2).
FT VARSPLIC 97 378 /FTID=VSP_005851.
FT VARSPLIC 97 378 Missing (in isoform 2).
FT VARSPLIC 97 378 /FTID=VSP_005852.
SQ SEQUENCE 378 AA; 43722 MW; E669C31BCA365AA0 CRC64;

Query Match 90.5%; Score 76; DB 1; Length 378;
Best Local Similarity 92.9%; Pred. No. 0.36;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
Db 241 GGGGGGGGGGGGGG 254

RESULT 9
HXB3 MOUSE STANDARD; PRT; 433 AA.
ID P09076; P10285; Q61680;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-B3 (Hox-2.7) (MH-23).
GN HOBX3 OR HOXB-3 OR HOX-2.7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92258392; PubMed=1592411;
RA Sham M.H., Hunt P., Nonchev S., Papalopolu N., Graham A.,
RA Boncinelli E., Krumlauf R.;
RT "Analysis of the murine Hox-2.7 gene: conserved alternative
transcripts with differential distributions in the nervous system and
the potential for shared regulatory regions.";
RL EMBO J. 11:1825-1836(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95196953; PubMed=7890121;
RA Brown W.M., Taylor G.R.;
RT "The 5'-sequence of the murine Hox-b3 (Hox-2.7) gene and its intron
contain multiple transcription-regulatory elements.";
RL Int. J. Biochem. 26:1403-1409(1994).
RN [3]
RP SEQUENCE OF 152-361 FROM N.A.
RX MEDLINE=88054465; PubMed=2830503;
RA Lonai P., Arman E., Czesnek H., Ruddle F.H., Blatt C.;
RT "New murine homeoboxes: structure, chromosomal assignment, and
differential expression in adult erythropoiesis.";
[4]
DNA 6:409-418(1987).
RP SEQUENCE OF 181-265 FROM N.A.
RX MEDLINE=89091992; PubMed=2463210;
RA Graham A., Papalopolu N., Lorimer J., McVey J.H., Tuddenham E.G.D.,
RA Krumlauf R.;
RT "Characterization of a murine homeo box gene, Hox-2.6, related to the
Drosophila Deformed Gene";
RL Genes Dev. 2:1424-1438(1988).
CC -!- FUNCTION: Sequence-specific transcription factor which is part of
a developmental regulatory system that provides cells with
specific positional identities on the anterior-posterior axis.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the Antp homeobox family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; X66177; CAA46951.1; --
DR EMBL; U02278; AAB60496.1; --
DR EMBL; M18168; AAA37840.1; --
DR PIR; S20963; S20963.
DR HSP; P02833; ISAN.
DR TRANSFAC; T01724; --
DR MGD; MGI:96184; HoxB3.
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00025; ANTENNAPEDIA.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00032; ANTENNAPEDIA; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT SITE 129 134 ANTP-TYPE HEXAPEPTIDE.
FT DOMAIN 154 181 GLY-RICH.
FT DOMAIN 191 250 HOMEBOX.
FT CONFLICT 113 113 G -> C (IN REF. 1).
FT CONFLICT 119 119 A -> S (IN REF. 1).
FT CONFLICT 132 168 GCGGGGGGGGGGGGGG -> RLWWRPAVAAAAAVR
(IN REF. 3).
FT CONFLICT 182 182 D -> N (IN REF. 4).
FT CONFLICT 216 217 LC -> FV (IN REF. 3).
FT CONFLICT 330 330 S -> L (IN REF. 3).
FT CONFLICT 342 361 GAYGPTMGSGSPVYGGGY -> APTGPPEPCVRCMWAG
VAT (IN REF. 3).
SQ SEQUENCE 433 AA; 44353 MW; 9AD3C922663612A6 CRC64;

Query Match 90.5%; Score 76; DB 1; Length 433;
Best Local Similarity 86.7%; Pred. No. 0.41;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
Db 156 GGGGGGGGGGGGGG 170

RESULT 10
DOP2 DROME
ID DOP2 DROME STANDARD; PRT; 539 AA.
AC Q24563; Q24569; Q9VAJ8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

DE Dopamine receptor 2.
GN DOPR2 OR DOPR99B OR DAMB OR CG18741.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM LONG).
RP STRAIN=Canton-S; TISSUE=Head;
RX MEDLINE=96242088; PubMed=8656286;
RA Feng G., Hannan F., Reale V., Hon Y.Y., Kousky C.T., Evans P.D.,
RA Hall L.M.;
RT "Cloning and functional characterization of a novel dopamine receptor
RT from Drosophila melanogaster";
RL J. Neurosci. 16:3925-3933(1996).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM SHORT).
RP STRAIN=Canton-S;
RX MEDLINE=96259265; PubMed=8663989;
RA Han K.-A., Millar N.S., Grotewiel M.S., Davis R.L.;
RT "DAMB, a novel dopamine receptor expressed specifically in Drosophila
RT mushroom bodies";
RL Neuron 16:1127-1135(1996).
RN [3]
RN SEQUENCE FROM N.A. (ISOFORM SHORT).
RP STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhargava A., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fodor C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Receptor for dopamine. The activity of this receptor is
CC mediated by G proteins which activate adenylyl cyclase. Also
CC capable of generating a calcium signal. In terms of antagonist
CC responses, would be classed with the D1-like dopamine receptor
CC group. This receptor an attractive candidate for initiating
CC biochemical cascades underlying olfactory learning.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q24563-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q24563-2; Sequence=VSP_001877;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Expressed in both central and peripheral
CC nervous systems.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U34383; AAC47161.1; -;
CC EMBL; U61264; AAB08000.1; -;
CC EMBL; AB003770; AAF56908.2; -;
CC FlyBase; FBgn0015129; DOPR2.
CC GO; GO:0004952; F:dopamine receptor activity; IDA.
CC InterPro; IPR000276; GPCR_Rhodopsn.
CC Pfam; PF00001; 7tm.1; 1.
CC PRINTS; PR00237; GPCRHOPOPSN.
CC PRINTS; PR00960; LMBPPOTEIN.
CC PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
CC PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
CC KW G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Phosphorylation; Lipoprotein; Palmitate; Alternative splicing.
CC DOMAIN 1 113
CC TRANSMEM 114 134
CC DOMAIN 135 145
CC TRANSMEM 146 166
CC DOMAIN 167 189
CC TRANSMEM 190 206
CC DOMAIN 207 227
CC TRANSMEM 228 248
CC DOMAIN 249 266
CC TRANSMEM 267 287
CC DOMAIN 288 420
CC TRANSMEM 421 441
CC DOMAIN 442 453
CC TRANSMEM 454 474
CC DOMAIN 475 539
CC TRANSMEM 539 559
CC DOMAIN 560 580
CC TRANSMEM 581 601
CC DOMAIN 602 622
CC TRANSMEM 623 643
CC DOMAIN 644 664
CC TRANSMEM 665 685
CC DOMAIN 686 706
CC TRANSMEM 707 727
CC DOMAIN 728 748
CC TRANSMEM 749 769
CC DOMAIN 770 790
CC TRANSMEM 791 811
CC DOMAIN 812 832
CC TRANSMEM 833 853
CC DOMAIN 854 874
CC TRANSMEM 875 895
CC DOMAIN 896 916
CC TRANSMEM 917 937
CC DOMAIN 938 958
CC TRANSMEM 959 979
CC DOMAIN 980 1000
CC TRANSMEM 1001 1021
CC DOMAIN 1022 1042
CC TRANSMEM 1043 1063
CC DOMAIN 1064 1084
CC TRANSMEM 1085 1105
CC DOMAIN 1106 1126
CC TRANSMEM 1127 1147
CC DOMAIN 1148 1168
CC TRANSMEM 1169 1189
CC DOMAIN 1190 1210
CC TRANSMEM 1211 1231
CC DOMAIN 1232 1252
CC TRANSMEM 1253 1273
CC DOMAIN 1274 1294
CC TRANSMEM 1295 1315
CC DOMAIN 1316 1336
CC TRANSMEM 1337 1357
CC DOMAIN 1358 1378
CC TRANSMEM 1379 1399
CC DOMAIN 1400 1420
CC TRANSMEM 1421 1441
CC DOMAIN 1442 1462
CC TRANSMEM 1463 1483
CC DOMAIN 1484 1504
CC TRANSMEM 1505 1525
CC DOMAIN 1526 1546
CC TRANSMEM 1547 1567
CC DOMAIN 1568 1588
CC TRANSMEM 1589 1609
CC DOMAIN 1610 1630
CC TRANSMEM 1631 1651
CC DOMAIN 1652 1672
CC TRANSMEM 1673 1693
CC DOMAIN 1694 1714
CC TRANSMEM 1715 1735
CC DOMAIN 1736 1756
CC TRANSMEM 1757 1777
CC DOMAIN 1778 1798
CC TRANSMEM 1799 1819
CC DOMAIN 1820 1840
CC TRANSMEM 1841 1861
CC DOMAIN 1862 1882
CC TRANSMEM 1883 1903
CC DOMAIN 1904 1924
CC TRANSMEM 1925 1945
CC DOMAIN 1946 1966
CC TRANSMEM 1967 1987
CC DOMAIN 1988 2008
CC TRANSMEM 2009 2029
CC DOMAIN 2030 2050
CC TRANSMEM 2051 2071
CC DOMAIN 2072 2092
CC TRANSMEM 2093 2113
CC DOMAIN 2114 2134
CC TRANSMEM 2135 2155
CC DOMAIN 2156 2176
CC TRANSMEM 2177 2197
CC DOMAIN 2198 2218
CC TRANSMEM 2219 2239
CC DOMAIN 2240 2260
CC TRANSMEM 2261 2281
CC DOMAIN 2282 2302
CC TRANSMEM 2303 2323
CC DOMAIN 2324 2344
CC TRANSMEM 2345 2365
CC DOMAIN 2366 2386
CC TRANSMEM 2387 2407
CC DOMAIN 2408 2428
CC TRANSMEM 2429 2449
CC DOMAIN 2450 2470
CC TRANSMEM 2471 2491
CC DOMAIN 2492 2512
CC TRANSMEM 2513 2533
CC DOMAIN 2534 2554
CC TRANSMEM 2555 2575
CC DOMAIN 2576 2596
CC TRANSMEM 2597 2617
CC DOMAIN 2618 2638
CC TRANSMEM 2639 2659
CC DOMAIN 2660 2680
CC TRANSMEM 2681 2701
CC DOMAIN 2702 2722
CC TRANSMEM 2723 2743
CC DOMAIN 2744 2764
CC TRANSMEM 2765 2785
CC DOMAIN 2786 2806
CC TRANSMEM 2807 2827
CC DOMAIN 2828 2848
CC TRANSMEM 2849 2869
CC DOMAIN 2870 2890
CC TRANSMEM 2891 2911
CC DOMAIN 2912 2932
CC TRANSMEM 2933 2953
CC DOMAIN 2954 2974
CC TRANSMEM 2975 2995
CC DOMAIN 2996 3016
CC TRANSMEM 3017 3037
CC DOMAIN 3038 3058
CC TRANSMEM 3059 3079
CC DOMAIN 3080 3100
CC TRANSMEM 3101 3121
CC DOMAIN 3122 3142
CC TRANSMEM 3143 3163
CC DOMAIN 3164 3184
CC TRANSMEM 3185 3205
CC DOMAIN 3206 3226
CC TRANSMEM 3227 3247
CC DOMAIN 3248 3268
CC TRANSMEM 3269 3289
CC DOMAIN 3290 3310
CC TRANSMEM 3311 3331
CC DOMAIN 3332 3352
CC TRANSMEM 3353 3373
CC DOMAIN 3374 3394
CC TRANSMEM 3395 3415
CC DOMAIN 3416 3436
CC TRANSMEM 3437 3457
CC DOMAIN 3458 3478
CC TRANSMEM 3479 3499
CC DOMAIN 3500 3520
CC TRANSMEM 3521 3541
CC DOMAIN 3542 3562
CC TRANSMEM 3563 3583
CC DOMAIN 3584 3604
CC TRANSMEM 3605 3625
CC DOMAIN 3626 3646
CC TRANSMEM 3647 3667
CC DOMAIN 3668 3688
CC TRANSMEM 3689 3709
CC DOMAIN 3710 3730
CC TRANSMEM 3731 3751
CC DOMAIN 3752 3772
CC TRANSMEM 3773 3793
CC DOMAIN 3794 3814
CC TRANSMEM 3815 3835
CC DOMAIN 3836 3856
CC TRANSMEM 3857 3877
CC DOMAIN 3878 3898
CC TRANSMEM 3899 3919
CC DOMAIN 3920 3940
CC TRANSMEM 3941 3961
CC DOMAIN 3962 3982
CC TRANSMEM 3983 4003
CC DOMAIN 4004 4024
CC TRANSMEM 4025 4045
CC DOMAIN 4046 4066
CC TRANSMEM 4067 4087
CC DOMAIN 4088 4108
CC TRANSMEM 4109 4129
CC DOMAIN 4130 4150
CC TRANSMEM 4151 4171
CC DOMAIN 4172 4192
CC TRANSMEM 4193 4213
CC DOMAIN 4214 4234
CC TRANSMEM 4235 4255
CC DOMAIN 4256 4276
CC TRANSMEM 4277 4297
CC DOMAIN 4298 4318
CC TRANSMEM 4319 4339
CC DOMAIN 4340 4360
CC TRANSMEM 4361 4381
CC DOMAIN 4382 4402
CC TRANSMEM 4403 4423
CC DOMAIN 4424 4444
CC TRANSMEM 4445 4465
CC DOMAIN 4466 4486
CC TRANSMEM 4487 4507
CC DOMAIN 4508 4528
CC TRANSMEM 4529 4549
CC DOMAIN 4550 4570
CC TRANSMEM 4571 4591
CC DOMAIN 4592 4612
CC TRANSMEM 4613 4633
CC DOMAIN 4634 4654
CC TRANSMEM 4655 4675
CC DOMAIN 4676 4696
CC TRANSMEM 4697 4717
CC DOMAIN 4718 4738
CC TRANSMEM 4739 4759
CC DOMAIN 4760 4780
CC TRANSMEM 4781 4801
CC DOMAIN 4802 4822
CC TRANSMEM 4823 4843
CC DOMAIN 4844 4864
CC TRANSMEM 4865 4885
CC DOMAIN 4886 4906
CC TRANSMEM 4907 4927
CC DOMAIN 4928 4948
CC TRANSMEM 4949 4969
CC DOMAIN 4970 4990
CC TRANSMEM 4991 5011
CC DOMAIN 5012 5032
CC TRANSMEM 5033 5053
CC DOMAIN 5054 5074
CC TRANSMEM 5075 5095
CC DOMAIN 5096 5116
CC TRANSMEM 5117 5137
CC DOMAIN 5138 5158
CC TRANSMEM 5159 5179
CC DOMAIN 5180 5200
CC TRANSMEM 5201 5221
CC DOMAIN 5222 5242
CC TRANSMEM 5243 5263
CC DOMAIN 5264 5284
CC TRANSMEM 5285 5305
CC DOMAIN 5306 5326
CC TRANSMEM 5327 5347
CC DOMAIN 5348 5368
CC TRANSMEM 5369 5389
CC DOMAIN 5390 5410
CC TRANSMEM 5411 5431
CC DOMAIN 5432 5452
CC TRANSMEM 5453 5473
CC DOMAIN 5474 5494
CC TRANSMEM 5495 5515
CC DOMAIN 5516 5536
CC TRANSMEM 5537 5557
CC DOMAIN 5558 5578
CC TRANSMEM 5579 5599
CC DOMAIN 5600 5620
CC TRANSMEM 5621 5641
CC DOMAIN 5642 5662
CC TRANSMEM 5663 5683
CC DOMAIN 5684 5704
CC TRANSMEM 5705 5725
CC DOMAIN 5726 5746
CC TRANSMEM 5747 5767
CC DOMAIN 5768 5788
CC TRANSMEM 5789 5809
CC DOMAIN 5810 5830
CC TRANSMEM 5831 5851
CC DOMAIN 5852 5872
CC TRANSMEM 5873 5893
CC DOMAIN 5894 5914
CC TRANSMEM 5915 5935
CC DOMAIN 5936 5956
CC TRANSMEM 5957 5977
CC DOMAIN 5978 5998
CC TRANSMEM 5999 6019
CC DOMAIN 6020 6040
CC TRANSMEM 6041 6061
CC DOMAIN 6062 6082
CC TRANSMEM 6083 6103
CC DOMAIN 6104 6124
CC TRANSMEM 6125 6145
CC DOMAIN 6146 6166
CC TRANSMEM 6167 6187
CC DOMAIN 6188 6208
CC TRANSMEM 6209 6229
CC DOMAIN 6230 6250
CC TRANSMEM 6251 6271
CC DOMAIN 6272 6292
CC TRANSMEM 6293 6312
CC DOMAIN 6313 6333
CC TRANSMEM 6334 6354
CC DOMAIN 6355 6375
CC TRANSMEM 6376 6396
CC DOMAIN 6397 6416
CC TRANSMEM 6417 6437
CC DOMAIN 6438 6458
CC TRANSMEM 6459 6479
CC DOMAIN 6480 6500
CC TRANSMEM 6501 6521
CC DOMAIN 6522 6542
CC TRANSMEM 6543 6563
CC DOMAIN 6564 6584
CC TRANSMEM 6585 6605
CC DOMAIN 6606 6626
CC TRANSMEM 6627 6647
CC DOMAIN 6648 6668
CC TRANSMEM 6669 6689
CC DOMAIN 6690 6710
CC TRANSMEM 6711 6731
CC DOMAIN 6732 6752
CC TRANSMEM 6753 6773
CC DOMAIN 6774 6794
CC TRANSMEM 6795 6815
CC DOMAIN 6816 6836
CC TRANSMEM 6837 6857
CC DOMAIN 6858 6878
CC TRANSMEM 6879 6899
CC DOMAIN 6900 6920
CC TRANSMEM 6921 6941
CC DOMAIN 6942 6962
CC TRANSMEM 6963 6983
CC DOMAIN 6984 7004
CC TRANSMEM 7005 7025
CC DOMAIN 7026 7046
CC TRANSMEM 7047 7067
CC DOMAIN 7068 7088
CC TRANSMEM 7089 7109
CC DOMAIN 7110 7130
CC TRANSMEM 7131 7151
CC DOMAIN 7152 7172
CC TRANSMEM 7173 7193
CC DOMAIN 7194 7214
CC TRANSMEM 7215 7235
CC DOMAIN 7236 7256
CC TRANSMEM 7257 7277
CC DOMAIN 7278 7298
CC TRANSMEM 7299 7319
CC DOMAIN 7320 7340
CC TRANSMEM 7341 7361
CC DOMAIN 7362 7382
CC TRANSMEM 7383 7403
CC DOMAIN 7404 7424
CC TRANSMEM 7425 7445
CC DOMAIN 7446 7466
CC TRANSMEM 7467 7487
CC DOMAIN 7488 7508
CC TRANSMEM 7509 7529
CC DOMAIN 7530 7550
CC TRANSMEM 7551 7571
CC DOMAIN 7572 7592
CC TRANSMEM 7593 7613
CC DOMAIN 7614 7634
CC TRANSMEM 7635 7655
CC DOMAIN 7656 7676
CC TRANSMEM 7677 7697
CC DOMAIN 7698 7718
CC TRANSMEM 7719 7739
CC DOMAIN 7740 7760
CC TRANSMEM 7761 7781
CC DOMAIN 7782 7802
CC TRANSMEM 7803 7823
CC DOMAIN 7824 7844
CC TRANSMEM 7845 7865
CC DOMAIN 7866 7886
CC TRANSMEM 7887 7907
CC DOMAIN 7908 7928
CC TRANSMEM 7929 7949
CC DOMAIN 7950 7970
CC TRANSMEM 7971 7991
CC DOMAIN 7992 8012
CC TRANSMEM 8013 8033
CC DOMAIN 8034 8054
CC TRANSMEM 8055 8075
CC DOMAIN 8076 8096
CC TRANSMEM 8097 8117
CC DOMAIN 8118 8138
CC TRANSMEM 8139 8159
CC DOMAIN 8160 8180
CC TRANSMEM 8181 8201
CC DOMAIN 8202 8222
CC TRANSMEM 8223 8243
CC DOMAIN 8244 8264
CC TRANSMEM 8265 8285
CC DOMAIN 8286 8306
CC TRANSMEM 8307 8327
CC DOMAIN 8328 8348
CC TRANSMEM 8349 8369
CC DOMAIN 8370 8390
CC TRANSMEM 8391 8411
CC DOMAIN 8412 8432
CC TRANSMEM 8433 8453
CC DOMAIN 8454 8474
CC TRANSMEM 8475 8495
CC DOMAIN 8496 8516
CC TRANSMEM 8517 8537
CC DOMAIN 8538 8558
CC TRANSMEM 8559 8579
CC DOMAIN 8580 8600
CC TRANSMEM 8601 8621
CC DOMAIN 8622 8642
CC TRANSMEM 8643 8663
CC DOMAIN 8664 8684
CC TRANSMEM 8685 8705
CC DOMAIN 8706 8726
CC TRANSMEM 8727 8747
CC DOMAIN 8748 8768
CC TRANSMEM 8769 8789
CC DOMAIN 8790 8810
CC TRANSMEM 8811 8831
CC DOMAIN 8832 8852
CC TRANSMEM 8853 8873
CC DOMAIN 8874 8894
CC TRANSMEM 8895 8915
CC DOMAIN 8916 8936
CC TRANSMEM 8937 8957
CC DOMAIN 8958 8978
CC TRANSMEM 8979 8999
CC DOMAIN 9000 9020
CC TRANSMEM 9021 9041
CC DOMAIN 9042 9062
CC TRANSMEM 9063 9083
CC DOMAIN 9084 9104
CC TRANSMEM 9105 9125
CC DOMAIN 9126 9146
CC TRANSMEM 9147 9167
CC DOMAIN 9168 9188
CC TRANSMEM 9189 9209
CC DOMAIN 9210 9230
CC TRANSMEM 9231 9251
CC DOMAIN 9252 9272
CC TRANSMEM 9273 9293
CC DOMAIN 9294 9314
CC TRANSMEM 9315 9335
CC DOMAIN 9336 9356
CC TRANSMEM 9357 9377
CC DOMAIN 9378 9398
CC TRANSMEM 9399 9419
CC DOMAIN 9420 9440
CC TRANSMEM 9441 9461
CC DOMAIN 9462 9482
CC TRANSMEM 9483 9503
CC DOMAIN 9504 9524
CC TRANSMEM 9525 9545
CC DOMAIN 9546 9566
CC TRANSMEM 9567 9587
CC DOMAIN 9588 9608
CC TRANSMEM 9609 9629
CC DOMAIN 9630 9650
CC TRANSMEM 9651 9671
CC DOMAIN 9672 9692
CC TRANSMEM 9693 9713
CC DOMAIN 9714 9734
CC TRANSMEM 9735 9755
CC DOMAIN 9756 9776
CC TRANSMEM 9777 9797
CC DOMAIN 9798 9818
CC TRANSMEM 9819 9839
CC DOMAIN 9840 9860
CC TRANSMEM 9861 9881
CC DOMAIN 9882 9902
CC TRANSMEM 9903 9923
CC DOMAIN 9924 9944
CC TRANSMEM 9945 9965
CC DOMAIN 9966 9986
CC TRANSMEM 9987 10000
CC SEQUENCE 539 AA; 59505 MW; 32FDDC0E935AF4B3 CRC64;
CC
CC Query Match 90.5%; Score 76; DB 1; Length 539;
CC Best Local Similarity 86.7%; Pred. No. 0.5;
CC Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 GGGGGGGGGGGGGGGG 15
CC |||||
CC Db 338 GGGGGGGGGGGGGGGG 352
CC
CC RESULT 11
CC T7L1_HUMAN

ID T7L1_HUMAN STANDARD; PRT; 588 AA.
 AC Q9HCS4; Q9NP00;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transcription factor 7-like 1 (HMG-box transcription factor 3) (TCF-3).
 DE TCF7L1 OR TCF3.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal lung;
 RX MEDLINE=20535962; PubMed=11085512;
 RA Sagara N., Katoh M.;
 RT "Mitomycin C resistance induced by TCF-3 overexpression in gastric cancer cell line MKN28 is associated with DT-diaphorase down-regulation.";
 RT Cancer Res. 60:5959-5962(2000).
 RN [2]
 RP SEQUENCE OF 331-419 FROM N.A.
 RX MEDLINE=92158676; PubMed=1741298;
 RA Castrop J., van Norren K., Clevers H.C.;
 RT "A gene family of HMG-box transcription factors with homology to TCF-1.";
 RT Nucleic Acids Res. 20:611-611(1992).
 RN [3]
 RP TISSUE-SPECIFICITY.
 RX MEDLINE=99113953; PubMed=9916915;
 RA Barker N., Huels G., Korinek V., Clevers H.;
 RT "Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium.";
 RT Am. J. Pathol. 154:29-35(1999).
 CC -I- FUNCTION: Participates in the Wnt signaling pathway. Binds to DNA and acts as repressor in the absence of CTNNB1, and as activator in its presence. Necessary for the terminal differentiation of epidermal cells, the formation of keratohyalin granules and the development of the barrier function of the epidermis (By similarity). Down-regulates NQO1, leading to increased mitomycin C resistance.
 CC -I- SUBUNIT: Binds the armadillo repeat of CTNNB1 and forms a stable complex (By similarity).
 CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- TISSUE SPECIFICITY: Detected in hair follicles and skin keratinocytes, and at lower levels in stomach epithelium.
 CC -I- DOMAIN: The putative Grncho interaction domain between the N-terminal CTNNB1 binding domain and the HMG-box is necessary for repression of the transactivation mediated by TCF7L1 and CTNNB1 (By similarity).
 CC -I- SIMILARITY: Belongs to the TCF/LEF family.
 CC -I- SIMILARITY: Contains 1 HMG box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB031046; BAB18185.1; -;
 CC EMBL; X62870; CAB91064.1; -;
 CC HSSP; P27782; 2LEF.
 CC Genew; HGNC:11640; TCF7L1.
 CC MTM; 604652; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003700; F:transcription factor activity; NAS.
 CC GO; GO:0006325; P:establishment and/or maintenance of chromatin...; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC GO; GO:0030111; P:regulation of Wnt receptor signaling pathway; NAS.
 CC InterPro; IPR000910; HMG_12_box.

DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS01118; HMG_BOX_2; 1.
 KW Transcription regulation; Activator; Repressor; Trans-acting factor;
 KW Nuclear protein; DNA-binding; Wnt signaling pathway.
 FT DOMAIN 1 74 CTNNB1-BINDING (BY SIMILARITY).
 FT DNA_BIND 346 414 HMG_BOX.
 FT DOMAIN 421 427 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 5 29 GLY-RICH.
 FT DOMAIN 117 326 PRO-RICH.
 SQ SEQUENCE 588 AA; 82FB0C9300482A02 CRC64;
 Query Match 90.5%; Score 76; DB 1; Length 588;
 Best Local Similarity 92.3%; Pred. No. 0.54;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GGGGGGGGGGGGGG 14
 DB 6 GGGGGGGGGGGGGG 19
 RESULT 12
 SR68_CANFA STANDARD; PRT; 622 AA.
 ID SR68_CANFA
 AC Q00004;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Signal recognition particle 68 kDa protein (SRP68).
 GN SRP68.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=91092392; PubMed=1702390;
 RA Herz J., Flint N., Stanley K., Frank R., Dobberstein B.;
 RT "The 68 kDa protein of signal recognition particle contains a glycine-rich region also found in certain RNA-binding proteins.";
 RT FEBS Lett. 276:103-107(1990).
 CC -I- FUNCTION: Signal-recognition-particle assembly has a crucial role in targeting secretory proteins to the rough endoplasmic reticulum membrane. SRP68 binds the 7S RNA, SRP72 binds to this complex subsequently. This ribonucleoprotein complex might interact directly with the docking protein in the ER membrane and possibly participate in the elongation arrest function.
 CC -I- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54, SRP19, SRP14 and SRP9.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic and nuclear; nucleolar (By similarity).
 CC -I- MISCELLANEOUS: The RNA binding domain is located near the N-terminus.
 CC -I- SIMILARITY: Belongs to the SRP68 family.
 CC -I- CAUTION: Some authors found genomic clones that have 9 or 12 consecutive glycine residues instead of 15 (AA 9-27).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X53744; CAA37773.1; ALT_SEQ.
 CC PIR; A58947; A58947.
 CC InterPro; IPR008941; TPR-like.
 KW Signal recognition particle; Ribonucleoprotein; RNA-binding;
 KW Nuclear protein.
 FT DOMAIN 9 27 POLY-GLY.
 SQ SEQUENCE 622 AA; 70275 MW; DB03DFE0DAE8B942 CRC64;

Query Match 90.5%; Score 76; DB 1; Length 622;
 Best Local Similarity 86.7%; Pred. No. 0.57;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGS 15
 DB 10 GGGGGGGGGGGGGS 24

RESULT 13

ID DUS8 MOUSE STANDARD; PRT; 663 AA.
 AC 009112;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (Neuronal tyrosine threonine phosphatase 1).
 GN DUSP8 OR NTRP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96311565; PubMed=8733137;
 RA Theodosiou A.M., Rodrigues N.R., Nesbit M.A., Ambrose H.J.,
 RA Paterson H., McLellan-Arnold E., Boyd Y., Leversha M.A., Owen N.,
 RA Blake D.J., Ashworth A., Davies K.E.;
 RT "A member of the MAP kinase phosphatase gene family in mouse
 RT containing a complex trinucleotide repeat in the coding region.";
 RL Hum. Mol. Genet. 5:675-684(1996).
 CC -!- FUNCTION: This protein shows both activity toward tyrosine-protein
 CC phosphate as well as with serine/threonine-protein phosphate (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -!- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in brain and lung.
 CC -!- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
 CC Non-receptor class dual specificity subfamily.
 CC -!- SIMILARITY: Contains 1 rhodanese domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; X95518; CAA64772.1; -;
 HSSP; Q16828; IMKP.
 DR MGD; Q16828; Dusp8.
 DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK phosph.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR Pfam; PF00782; DSPC; 1.
 DR PRINTS; PR01764; MAPKPHPTASE.
 DR SMART; SM00195; DSPC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS02026; RHODANES 3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE 1; 1.
 DR PROSITE; PS00056; TYR_PHOSPHATASE 2; 1.
 DR PROSITE; PS00054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolase; Nuclear protein.
 KW DOMAIN 23 138 RHODANES.
 FT DOMAIN 162 432 PROTEIN-TYROSINE PHOSPHATASE.

FT DOMAIN 452 459 POLY-ARG.
 FT DOMAIN 555 558 POLY-SER.
 FT DOMAIN 559 576 POLY-GLY.
 FT DOMAIN 577 600 POLY-SER.
 FT DOMAIN 311 552 PRO-RICH.
 FT ACT_SITE 246 246 PHOSPHOCYCSTEINE INTERMEDIATE (BY
 FT SIMILARITY).
 SQ SEQUENCE 663 AA; 68847 MW; 416F429A12C1FA7C CRC64;

Query Match 90.5%; Score 76; DB 1; Length 663;
 Best Local Similarity 86.7%; Pred. No. 0.6;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGS 15
 DB 563 GGGGGGGGGGGGGS 577

RESULT 14

ID BOMD MOUSE STANDARD; PRT; 688 AA.
 AC 054839; Q9QYG7;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bomesodermin homolog.
 GN BOMES OR TBR2
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99337662; PubMed=10407135;
 RA Kimura N., Nakashima K., Ueno M., Taga T.;
 RT "A novel mammalian T-box-containing gene, Tbr2, expressed in mouse
 RT developing brain.";
 RL Brain Res. Dev. Brain Res. 115:183-193(1999).
 RN [2]
 RP SEQUENCE OF 278-457 FROM N.A.
 RX MEDLINE=98163742; PubMed=9503012;
 RA Wattler S., Russ A., Evans M., Nehls M.;
 RT "A combined analysis of genomic and primary protein structure defines
 RT the phylogenetic relationship of new members of the T-box family.";
 RL Genomics 48:24-33(1998).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: Contains 1 T-box domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AB031037; BAA83416.1; -;
 DR EMBL; AF013281; AAC16233.1; -;
 DR HSSP; P24781; IXBR.
 DR MGD; MGI:1201683; Bomes.
 DR InterPro; IPR008967; P53-like.
 DR InterPro; IPR001699; TF_T-box.
 DR Pfam; PF00907; T-box; 1.
 DR PRINTS; PR00937; TBOX.
 DR SMART; SM00425; TBOX; 1.
 DR PROSITE; PS01283; TBOX 1; 1.
 DR PROSITE; PS01264; TBOX 2; 1.
 DR PROSITE; PS0252; TBOX_3; 1.
 KW Developmental protein; Transcription regulation; DNA-binding;
 KW Nuclear protein.
 KW DOMAIN 27 41 POLY-GLY.
 FT DOMAIN 278 458 T-BOX.

```

FT DOMAIN 383 386 POLY-ASN.
SQ SEQUENCE 688 AA; 72638 MW; 197808989E920B92B CRC64;

Query Match 90.5%; Score 76; DB 1; Length 688;
Best Local Similarity 86.7%; Pred. No. 0.62;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
   |||||
Db 28 GGGGGGGGGGGGGG 42

RESULT 15
RFX1_HUMAN
ID RFX1_HUMAN STANDARD; PRT; 979 AA.
AC P22670;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MHC class II regulatory factor RFX1 (RFX) (Enhancer factor C) (EF-C).
GN RFX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071581; PubMed=22538877;
RA Reith W., Sanchez-Herrero C., Kober M., Silacci P., Mach C.,
RA Barras E., Mach B.;
RT "MHC class II regulatory factor RFX has a novel DNA-binding domain
and a functionally independent dimerization domain.";
RL Genes Dev. 4:1528-1540(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP IDENTITY BETWEEN RFX1 AND EF-C.
RX MEDLINE=94019311; PubMed=8413236;
RA Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA Reith W.;
RT "RFX1 is identical to enhancer factor C and functions as a
transactivator of the hepatitis B virus enhancer.";
RL Mol. Cell. Biol. 13:6375-6384(1993).
RN [4]
RP BINDING TO RPL30 PROMOTER.
RX MEDLINE=94040774; PubMed=8224874;
RA Safran G., Perry R.P.;
RT "Transcription factor RFX1 helps control the promoter of the mouse
ribosomal protein-encoding gene rpl30 by binding to its alpha
element.";
RL Gene 132:279-283(1993).

```

```

[5]
RN SHOWS THAT BLS II IS NOT DUE TO RFX1.
RX MEDLINE=92375076; PubMed=1508204;
RA Sanchez-Herrero C., Reith W., Silacci P., Mach B.;
RT "The DNA-binding defect observed in major histocompatibility complex
class II regulatory mutants concerns only one member of a family of
complexes binding to the X boxes of class II promoters.";
RL Mol. Cell. Biol. 12:4076-4083(1992).
CC -|- FUNCTION: Regulatory factor essential for MHC class II genes
expression. Binds to the X boxes of MHC class II genes. Also binds
to an inverted repeat (ENH1) required for hepatitis B virus genes
expression and to the most upstream element (alpha) of the RPL30
promoter.
CC -|- SUBUNIT: Binds DNA as a homodimer.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- SIMILARITY: Belongs to the RFX family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X58964; CAA41730.1; -
CC EMBL; A20498; CAA01506.1; -
CC EMBL; BC049826; AAH49826.1; -
CC PIR; A35913; A35913
CC PDB; 1DP7; 06-MAR-00.
CC TRANSFAC; T01673; -
CC GENEM; HGNC:9982; RFX1.
CC MIM; 600006; -
CC GO; GO:0003705; P:RNA polymerase II transcription factor acti...; TAS.
CC GO; GO:0006955; P:immune response; TAS.
CC InterPro; IPR007668; RFX1 trans act.
CC InterPro; IPR003150; RFX DNA binding.
CC Pfam; PF04589; RFX1 trans act; 1.
CC Pfam; PF02257; RFX DNA binding; 1.
CC DNA-binding: Transcription regulation; Activator; Nuclear protein;
KW 3D-structure.
FT DOMAIN 381 411 GLY-RICH
FT DNA_BIND 438 528 EXPERIMENTALLY DEDUCED.
FT DOMAIN 920 936 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 744 979 NECESSARY FOR DIMERIZATION.
SQ SEQUENCE 979 AA; 104728 MW; 556151F88C6AC9A2 CRC64;

Query Match 90.5%; Score 76; DB 1; Length 979;
Best Local Similarity 92.9%; Pred. No. 0.84;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
   |||||
Db 381 GGGGGGGGGGGGGG 394

Search completed: April 20, 2004, 10:25:36
Job time : 14.6 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:22:14 ; Search time 16.2 Seconds
(without alignments)
89.066 Million cell updates/sec

Title: US-08-930-480A-5

Perfect score: 84

Sequence: 1 GGGSGGGSGGGGS 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	95.2	592	2	endo-1,4-beta-gluc
2	79	94.0	104	2	holotricin 3 precu
3	78	92.9	249	2	single chain Fv an
4	77	91.7	206	2	gene M-twist prote
5	77	91.7	268	2	Ig heavy chain V r
6	76	90.5	80	2	hypothetical prote
7	76	90.5	165	1	glycine-rich cell
8	76	90.5	207	2	glycine-rich prote
9	76	90.5	221	2	glycine-rich cell
10	76	90.5	255	2	hypothetical prote
11	76	90.5	280	2	chitinase (EC 3.2
12	76	90.5	322	2	single-stranded-DN
13	76	90.5	378	2	U1 snRNP 70K prote
14	76	90.5	433	2	homeotic protein H
15	76	90.5	528	2	fus-like protein -
16	76	90.5	979	2	regulatory factor
17	76	90.5	1969	2	histidine kinase h
18	76	90.5	2783	1	alpha-fetoprotein
19	75	89.3	302	2	hypothetical prote
20	75	89.3	322	2	hypothetical prote
21	75	89.3	1226	2	calpain (EC 3.4.22
22	74	88.1	266	1	transcription fact
23	74	88.1	440	2	coat protein VP1 -
24	74	88.1	722	1	hypothetical prote
25	74	88.1	877	2	zinc finger bindin
26	74	88.1	895	2	bone morphogenetic
27	73	86.9	151	2	CBP20 preproprotei
28	73	86.9	211	2	calpain (EC 3.4.22
29	73	86.9	263	2	calpain (EC 3.4.22

30	73	86.9	266	1	CIPGL
31	73	86.9	268	1	CIHUL
32	73	86.9	272	2	T02745
33	73	86.9	291	1	S31415
34	73	86.9	333	2	A39065
35	73	86.9	367	2	UC6087
36	73	86.9	377	2	T04213
37	73	86.9	396	2	T49109
38	73	86.9	495	1	S31223
39	73	86.9	681	2	AB2155
40	73	86.9	901	2	UC6093
41	73	86.9	1473	2	T13855
42	72	85.7	106	2	F84797
43	72	85.7	113	2	S44750
44	72	85.7	136	2	T29282
45	72	85.7	188	2	S49192

ALIGNMENTS

RESULT 1

E82759
endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)
C/Species: Xylella fastidiosa
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C/Accession: E82759
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A/Reference number: A82515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: E82759
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-592 <SIN>
A/Cross-references: GB:AE003921; GB:AE003849; NID:99105710; PIDN:AAF83628.1; GSPDB:GN001.
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A/Reference number: A59328
A/Contents: annotation
C/Genetics:
A/Gene: XF0818

Query Match 95.2%; Score 80; DB 2; Length 592;

Best Local Similarity 100.0%; Pred. No. 0.29;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGSGGGSGGGG 14

Db 467 GGGSGGGSGGGG 480

RESULT 2

JC4190

holotricin 3 precursor - Holotrichia diomphalia

N/Alternate names: antifungal protein

C/Species: Holotrichia diomphalia

C/Date: 04-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999

C/Accession: JC4190

R/Lee, S.Y.; Moon, H.J.; Kurata, S.; Natori, S.; Lee, B.L.

Biol. Pharm. Bull. 18, 1049-1052, 1995

A;Title: Purification and cDNA cloning of an antifungal protein from the hemolymph of Hd
A;Reference number: JC4190; MUID:96073722; PMID:8535393
A;Accession: JC4190
A;Molecule type: mRNA
A;Residues: 1-104 <LEE>
A;Cross-references: DBJ:DL3744; NID:gl088433; PIDN:BAA02889.1; PID:dl003394; PID:gi7861
C;Species: Mus musculus (house mouse)
C;Comment: This protein is a Gly- and His-rich protein and a constitutive protein of lar
C;Keywords: hemolymph
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-104/Product: holotricin 3 #status predicted <MAT>

Query Match 94.0%; Score 79; DB 2; Length 104;
Best Local Similarity 93.3%; Pred. No. 0.08;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
|||||
Db 64 GGGGGGGGGGGGGG 78

RESULT 3
S41374
single chain Fv antibody - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C;Accession: S41374
R;Artsenko, O.; Weiler, E.W.; Muentz, K.; Conrad, U.
submitted to the EMBL Data Library, January 1994
A;Description: Construction and functional characterization of a single chain Fv antibod
A;Reference number: S41374
A;Accession: S41374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-249 <ART>
A;Cross-references: EMBL:Z29480

Query Match 92.9%; Score 78; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGGGGGGGGGGGG 15
|||||
Db 121 GGGGGGGGGGGGGG 134

RESULT 4
IS3066
gene M-twist protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: IS3066; I66795
R;Wolf, C.; Thiesse, C.; Stoetzel, C.; Thiesse, B.; Gerlinger, P.; Perrin-Schmitt, F.
Dev. Biol. 143, 363-373, 1991
A;Title: The M-twist gene of Mus is expressed in subsets of mesodermal cells and is clos
A;Reference number: IS3066; MUID:91122450; PMID:1840517
A;Accession: IS3066
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-206 <RES>
A;Cross-references: GB:M63649; NID:G202243; PIDN:AAA40514.1; PID:G202244
A;Accession: I66795
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-35, 'R', 37-90, 'P', 92-206 <RE2>
A;Cross-references: GB:M63650; NID:G202245; PIDN:AAA40515.1; PID:G202246
C;Genetics:
A;Gene: M-twist

Query Match 91.7%; Score 77; DB 2; Length 206;
Best Local Similarity 86.7%; Pred. No. 0.22;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15

Db 83 GGGGAGGGGGGGGG 97
|||||
|||||

RESULT 5

AS6446
Ig heavy chain V region (3H-3H scFv) - mouse (strain BALB/C)
C;Species: Mus musculus (house mouse)
C;Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 16-Aug-1996
C;Accession: AS6446
R;Tang, P.M.; Foltz, L.A.; Mahoney, W.C.; Schueler, P.A.
J. Biol. Chem. 270, 7829-7835, 1995
A;Title: A high affinity digoxin-binding protein displayed on M13 is functionally identifi
A;Reference number: AS6446; MUID:95229583; PMID:7713873
A;Accession: AS6446
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-268 <TAN>
A;Cross-references: GB:U20617
C;Keywords: heterotetramer; immunoglobulin

Query Match 91.7%; Score 77; DB 2; Length 268;
Best Local Similarity 93.3%; Pred. No. 0.28;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
|||||
Db 121 GGGGGGGGGGGGGG 135

RESULT 6

T10550
hypothetical protein T12G13.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T10550
R;Bevan, M.; Lemard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16533
A;Accession: T10550
A;Molecule type: DNA
A;Residues: 1-80 <BEV>
A;Cross-references: EMBL:AL080252; GSPDB:GN00062; ATSP:T12G13.70
A;Experimental source: cultivar Columbia; BAC clone T12G13
C;Genetics:
A;Gene: ATSP:T12G13.70
A;Map position: 4

Query Match 90.5%; Score 76; DB 2; Length 80;
Best Local Similarity 92.9%; Pred. No. 0.12;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
|||||
Db 22 GGGGGGGGGGGGGG 35

RESULT 7

KNRZG1
glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice
C;Species: Oryza sativa (rice)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C;Accession: S13385
R;Lei, M.; Wu, R.
Plant Mol. Biol. 16, 187-198, 1991
A;Title: A novel glycine-rich cell wall protein gene in rice.
A;Reference number: S13385; MUID:91370862; PMID:1716496
A;Accession: S13385
A;Molecule type: DNA
A;Residues: 1-165 <LEI>
A;Cross-references: EMBL:X53596; NID:G20246; PIDN:CAA37665.1; PID:G20247
C;Genetics:
A;Gene: grp-1

C:Superfamily: glycine-rich cell wall structural protein 1
 C:Keywords: cell wall; duplication; structural protein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-165/Product: Glycine-rich cell wall structural protein 1 #status predicted <MAT>
 F:30-55/Region: repeat R1
 F:56-62/Region: repeat R2
 F:62-92/Region: repeat R1
 F:93-99/Region: repeat R2
 F:100-131/Region: repeat R1
 F:132-138/Region: repeat R2
 F:139-160/Region: repeat R1

Query Match 90.5%; Score 76; DB 1; Length 165;
 Best Local Similarity 86.7%; Pred. No. 0.23;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 DB 111 GGGGGGGGGGGGGG 125

RESULT 8

T07381
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
 C:Accession: T07381
 R:Santino, C.G.; Stanford, G.L.; Conner, T.W.

A:Title: Developmental and transgenic analysis of two tomato fruit enhanced genes.
 A:Reference number: Z16000; MUID:97201476; PMID:9049262
 A:Accession: T07381
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-207 <SAN>
 A:Cross-references: EMBL:X95262; NID:g1166449; PIDN:CAA64559.1; PID:g1166450

A:Experimental source: cultivar UC82b; fruit

C:Genetics:

A:Gene: Tfm5

C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 90.5%; Score 76; DB 2; Length 207;
 Best Local Similarity 92.9%; Pred. No. 0.28;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
 |||||
 DB 97 GGGGGGGGGGGGGG 110

RESULT 9

T04592
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999

C:Accession: T04592
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.

A:Reference number: Z15378
 A:Accession: T04592

A:Molecule type: DNA

A:Residues: 1-221 <BEV>

A:Cross-references: EMBL:AL022141

A:Experimental source: cultivar Columbia; BAC clone F23E13

C:Genetics:

A:Map position: 4

A>Note: F23E13.120

Query Match 90.5%; Score 76; DB 2; Length 221;
 Best Local Similarity 86.7%; Pred. No. 0.3;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15

DB 168 GGGGGGGGGGGGGG 182
 |||||
 |||||

RESULT 10

B84777
 hypothetical protein At2g36120 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84777

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, J

eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84777

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-255 <STO>

A:Cross-references: GB:AE002093; NID:g4678224; PIDN:AAD26969.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36120

A:Map position: 2

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

Query Match 90.5%; Score 76; DB 2; Length 255;
 Best Local Similarity 92.9%; Pred. No. 0.34;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14
 |||||

DB 213 GGGGGGGGGGGGGG 226
 |||||

RESULT 11

A42424
 chitinase (EC 3.2.1.14) A - maize

C:Species: Zea mays (maize)

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999

C:Accession: A42424; A42260

R:Huynh, Q.K.; Hironaka, C.M.; Levine, E.B.; Smith, C.E.; Borgmeyer, J.R.; Shah, D.M.

J. Biol. Chem. 267, 6635-6640, 1992

A:Title: Antifungal proteins from plants. Purification, molecular cloning, and antifungal

A:Reference number: A42424; MUID:92202208; PMID:1551872

A:Accession: A42424

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-280 <HUY>

A:Experimental source: seed

A>Note: sequence inconsistent with nucleotide translation

R:Verburg, J.G.; Smith, C.E.; Lisek, C.A.; Huynh, Q.K.

J. Biol. Chem. 267, 3886-3893, 1992

A:Title: Identification of an essential tyrosine residue in the catalytic site of a chit-

o-1)-carboxylamide.

A:Reference number: A42260; MUID:92156129; PMID:1740436

A:Accession: A42260

A:Molecule type: protein

A:Residues: 180-195 <VER>

A>Note: the residue designated 'X' was determined to be derivatized tyrosine; therefore,

C:Superfamily: lectin-related plant chitinase; hevein chitin-binding domain homology; pl

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:26-61/Domain: hevein chitin-binding domain homology <HCB>

F:82-280/Domain: plant chitinase homology <PCH>

F:188/Active site: Tyr #status predicted

Query Match 90.5%; Score 76; DB 2; Length 280;
 Best Local Similarity 86.7%; Pred. No. 0.36;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||

Db 61 GGGGGGGGGGGGGG 75

RESULT 12

A45036
single-stranded-DNA-binding protein Pur alpha - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C;Accession: A45036
R;Bergemann, A.D.; Ma, Z.W.; Johnson, E.M.
Mol. Cell. Biol. 12, 5673-5682, 1992
A;Title: Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding protein Pur alpha - human
A;Reference number: A45036; MUID:93078769; PMID:1448097
A;Accession: A45036
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-322 <BER>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:119216)

Query Match 90.5%; Score 76; DB 2; Length 322;

Best Local Similarity 92.9%; Pred. No. 0.41;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14

Db 34 GGGGGGGGGGGGGG 47

RESULT 13

S04336
U1 snRNP 70K protein (long form) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 28-Feb-1990 #sequence_revision 30-Sep-1991 #text_change 24-Sep-1999
C;Accession: S04336
R;Hornig, H.; Fischer, U.; Costas, M.; Rauh, A.; Luehrmann, R.
Eur. J. Biochem. 182, 45-50, 1989
A;Title: Analysis of genomic clones of the murine U1RNP-associated 70-kDa protein reveal a family of related proteins
A;Reference number: S04336; MUID:89276388; PMID:2525092
A;Accession: S04336
A;Molecule type: DNA
A;Residues: 1-378 <HOR>
A;Cross-references: EMBL:X15769; NID:G55084; PIDN:CAA33777.1; PID:G763157
A;Note: the authors translated the codon GGC for residue 101 as Glu, ACT for residue 113 as 368 and 374 as Asp
C;Genetics:
A;Introns: 19/1; 40/3; 61/3; 89/1; 123/1; 152/2
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoproteins
C;Keywords: alternative splicing
F;34-101/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 90.5%; Score 76; DB 2; Length 378;

Best Local Similarity 92.9%; Pred. No. 0.47;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14

Db 241 GGGGGGGGGGGGGG 254

RESULT 14

S20963
homeotic protein Hox B3 - mouse
N;Alternate names: homeotic protein Hox 2.7
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 20-Aug-1999
C;Accession: S20963; D42694
R;Sham, M.H.; Hunt, P.; Nonchev, S.; Papalopulu, N.; Graham, A.; Boncinelli, E.; Krumlauf, J.B.
EMBO J. 11, 1825-1836, 1992
A;Title: Analysis of the murine Hox-2.7 gene: conserved alternative transcripts with different 5' ends
A;Reference number: S20963; MUID:92258392; PMID:1582411
A;Accession: S20963
A;Status: preliminary

A;Molecule type: mRNA
A;Residues: 1-433 <SHA>
A;Cross-references: GB:X66177; GB:S35628; GB:S35738; NID:g312229; PIDN:CAA46951.1; PID:g312229; GB:S35738; NID:g312229; PIDN:CAA46951.1; PID:g312229
R;Nazarali, A.; Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 2883-2887, 1992
A;Title: Hox-1.11 and Hox-4.9 homeobox genes.
A;Reference number: A42694; MUID:92212934; PMID:1348361
A;Accession: D42694
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 213-238 <NA>
A;Note: sequence extracted from NCBI backbone (NCBIN:92310, NCBIP:92316)
C;Superfamily: homeotic protein Hox B3; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;192-248/Domain: homeobox homology <HOX>

Query Match 90.5%; Score 76; DB 2; Length 433;

Best Local Similarity 86.7%; Pred. No. 0.53;

Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15

Db 156 GGGGGGGGGGGGGG 170

RESULT 15

G02127
fus-like protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 28-Jul-2003
C;Accession: G02127
R;Itoh, K.; Kawase, M.
submitted to the EMBL Data Library, September 1995
A;Reference number: G02127
A;Accession: G02127
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-528 <ITO>
A;Cross-references: EMBL:U36561; NID:g1040969; PIDN:AAA79948.1; PID:g1040970
C;Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology
F;289-364/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 90.5%; Score 76; DB 2; Length 528;

Best Local Similarity 92.9%; Pred. No. 0.63;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 14

Db 183 GGGGGGGGGGGGGG 196

Search completed: April 20, 2004, 10:27:41

Job time : 17.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:20:04 ; Search time 60.6 Seconds
(without alignments)
69.938 Million cell updates/sec

Title: US-08-930-480A-5

Perfect score: 84

Sequence: 1 GGGGGGGGGGGGS 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	100.0	15	2	AAR25983 Peptide m
2	84	100.0	15	2	AAR59500 Hydrophil
3	84	100.0	15	2	AAR85123 Gene deli
4	84	100.0	15	2	AAR76683 Human ONS
5	84	100.0	15	2	AAR99244 (Gly4Ser)
6	84	100.0	15	2	AAR95067 scFv, spac
7	84	100.0	15	2	AAR09323 Peptide 1
8	84	100.0	15	2	AAY49219 Sequence
9	84	100.0	15	2	Aaw10295 Peptide 1
10	84	100.0	15	2	Aaw35984 Peptide 1
11	84	100.0	15	2	Aaw87784 Antibody-
12	84	100.0	15	2	Aay43414 Peptide S
13	84	100.0	15	2	Aay33328 B6-srv pe
14	84	100.0	15	2	Aay03763 Linker pe
15	84	100.0	15	2	Aay21600 EP-919566
16	84	100.0	15	2	Aay27397 Flexible
17	84	100.0	15	2	Aa16564 Peptide 1
18	84	100.0	15	2	Aab29542 Linker pe
19	84	100.0	15	3	Aay99636 (gly4ser)
20	84	100.0	15	3	Aab22838 Single ch
21	84	100.0	15	3	Aay70606 Protein e
22	84	100.0	15	3	Aay79551 Linker pe
23	84	100.0	15	3	Aay79552 Linker pe
24	84	100.0	15	3	Aay90826 Linker am
25	84	100.0	15	3	Aab23816 Plaamid p

ALIGNMENTS

RESULT 1

AAR25983
ID AAR25983 standard; protein; 15 AA.
XX AAR25983;
AC AAR25983;
XX
DT 25-MAR-2003 (revised)
DT 21-JAN-1993 (first entry)
XX
DE Peptide monomer 21.
XX
KW Reverse peptide; microbial pathogen; phytotoxicity; head-to-tail;
KW proteolytic degradation; dimer; peptide bond; bridging group; omega loop.
XX
OS Synthetic.
XX
FN EP497366-A2.
XX
PD 05-AUG-1992.
XX
PF 31-JAN-1992; 92EP-00101616.
XX
PR 01-FEB-1991; 91US-00649784.
XX
PA (DONG) IST DONEGANI SPA GUIDO.
XX (ENTE) ENICHEM SPA.
PI Mapelli C, Dugas De Robertis C, Stahl GF, Bascomb NF;
PI Swardloff MD, Williams JI, Everett NP;
XX WPI; 1992-260816/32.
XX
PT Reverse antimicrobial peptide(s) and oligopeptide(s) - useful for
protecting plants from pathogens and for determining phytotoxicity.
XX Disclosure; Fig 1; 79pp; English.

The sequences given in AAR25983-83 are a collection of natural and reverse peptides which are active against at least one microbial pathogen and, preferably, at least one plant pathogen. It has been found that acceptable activity and acceptable levels of protection against at least one microbial pathogen and at least one microbial plant pathogen may be obtained by reversing the sequence of amino acids contained within naturally occurring antimicrobial peptides while maintaining the directionality of the peptide bonds. These peptides possess relatively low phytotoxicity and/or low susceptibility to proteolytic degradation. The oligopeptides may be used as dimers composed of two peptide units with or without an intervening bridge. The simplest structure taken by

CC these dimers is the "head-to-tail" configuration. This comprises at least
 CC one first peptide monomer and at least one second peptide monomer. Each
 CC peptide monomer has an N- and C-terminus, both of which are capable of
 CC forming peptide bonds. In the head-to-tail configuration the C-terminal
 CC amino acid of the first monomer peptide is directly bound to the N-
 CC terminus of the second monomer peptide, by a peptide bond, without an
 CC intervening bridging group. In other peptide dimers bridging groups may
 CC be used and may be as few as one amino acid but may be as large as 100
 CC amino acids in length and form omega loops or other secondary structures.
 CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
 CC correct PA field.)
 XX
 XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006; 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGSGGGSGGGSGGGGS 15
 |||||
 DB 1 GGGSGGGSGGGSGGGGS 15

RESULT 2

AAR59500
 ID AAR59500 standard; peptide; 15 AA.

XX
 AC AAR59500;

XX 25-MAR-2003 (revised)

DT 29-JUL-1994 (first entry)

XX Hydrophilic linker #1 to make single chain antibody.

XX Single chain antibody; sFv; heavy chain; light chain; variable domain;
 KW hydrophilic linker; antibodies.

XX Synthetic.

PH Key Location/Qualifiers

FT Region

1. .5 /note= "first of 3 repeat units"

XX W09402610-A1.

XX 03-FEB-1994.

XX 16-JUL-1993; 93WO-US006735.

XX 17-JUL-1992; 92US-00916939.

PR 17-MAR-1993; 93US-00045274.

XX (DAND) DANA FARBER CANCER INST INC.

PI Marasco WA, Haseltine WA;

DR WPI; 1994-048868/06.

XX Intracellular binding of antigens - by using antibody targetting with
 PT vector system, for e.g. tumour suppression.

XX Claim 35; Page 25; 155pp; English.

XX New vector systems comprise a sequence adapted for intracellular delivery
 CC and expression contg. a promoter operably linked to an antibody gene
 CC encoding an antibody which binds to a specific target antigen. The
 CC antibody is esp. a single chain antibody in which the heavy and light
 CC chain variable regions are joined via a hydrophilic linker peptide.
 CC Examples of suitable linkers are given in AAR59500- AAR59507, with
 CC AAR59500 being the most preferred linker. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX
 XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGSGGGSGGGSGGGGS 15
 |||||
 DB 1 GGGSGGGSGGGSGGGGS 15

RESULT 3

AAR85123
 ID AAR85123 standard; peptide; 15 AA.

XX
 AC AAR85123;

XX 06-JUN-1996 (first entry)

XX Gene delivery fusion protein flexon peptide.

XX Targeted nucleic acid; fusion protein; nucleic acid binding domain;
 KW gene delivery domain; cell; GAL4; interleukin; flexon; linker; primer;
 KW amplification; PCR; S.cerevisiae; gene therapy.

XX Synthetic.

XX W09528494-A1.

XX 26-OCT-1995.

XX 17-APR-1995; 95WO-US004738.

XX 15-APR-1994; 94US-00227858.

PR 19-OCT-1994; 94US-00326460.

XX (TARG-) TARGETED GENETICS CORP.

XX Overell RW, Weisser KE;

XX WPI; 1995-373808/48.

DR N-PSDB; AAT02970.

XX Fusion protein for delivering targetted nucleic acid to target cell -
 PT comprises a nucleic acid binding domain and a gene delivery domain, used
 PT in, e.g. gene therapy of Cystic fibrosis and in tumour vaccines.

XX Example 3; Page 49; 80pp; English.

XX A novel method of delivering a targetted nucleic acid involves a fusion
 CC protein comprising nucleic acid binding domain (NBD) linked to a gene
 CC delivery domain (GDD). The NBD binds the target DNA whilst the GDD
 CC mediates the delivery of the target DNA into the cell. An example of the
 CC fusion protein comprises the GAL4 NBD linked to the interleukin (IL)-2
 CC GDD. The NBD and GDD domain can be separated by a short flexible peptide
 CC linker termed a "flexon". The oligomers AAT02970-1 were annealed to
 CC encode such a "flexon". The annealed product was inserted between the
 CC coding sequence of the yeast GAL4 NBD and the IL-2 GDD in the plasmid
 CC PT3GAL4/IL-2m. This vector was transformed into E.coli DH10B for
 CC production of the fusion protein. The fusion protein has applications in
 CC gene therapy esp. for in vivo and in vitro gene delivery
 XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGSGGGSGGGSGGGGS 15
 |||||
 DB 1 GGGSGGGSGGGSGGGGS 15

RESULT 4


```

AAR76683
ID AAR76683 standard; protein; 15 AA.
XX
AC AAR76683;
XX
DT 18-JAN-1996 (first entry)
XX
DE Human ONS-M21 antibody Fv fragment linker peptide.
XX
KW Plasmid pSCFVT7-hm21; human; ONS-M21 antibody; chimeric protein;
KW medulloblastoma; brain tumour; treatment; diagnosis; Fv fragment.
XX
OS Synthetic.
XX
PN WO9514041-A1.
XX
PD 26-MAY-1995.
XX
PF 19-OCT-1994; 94WO-JP001763.
XX
PR 19-NOV-1993; 93JP-00291078.
XX
PA (CHUS ) CHUGAI SEIYAKU KK.
XX
PI Ohtomo T, Sato K, Tsuchiya M;
XX
DR WPI; 1995-200347/26.
XX
DR N-PSDB; AAQ94549.
XX
PT Reconstituted antibody against human medullo:blastoma cells - contains
PT high proportion of human antibody origin and has low antigenicity.
XX
XX
XX Claim 32; Page 103; 120pp; Japanese.
XX
CC AAQ94549 encodes AAR76683 a peptide linker, part of the human antibody
CC ONS-M21 Fv fragment. The fragment was used in the construction of a
CC human/murine chimeric antibody, reactive with human medullo- blastoma (a
CC brain tumour) cells. The chimeric antibody can be used in the diagnosis
CC and treatment of this disease
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
DB 1 GGGGGGGGGGGGGG 15
|||||
RESULT 5
AAR99244
ID AAR99244 standard; peptide; 15 AA.
XX
AC AAR99244;
XX
DT 28-NOV-1996 (first entry)
XX
DE (Gly4Ser)3 linker.
XX
KW Bioactive fusion protein; interleukin-12; IL-12; p35 subunit;
KW p40 subunit; antitumour; cytokine; tumour; melanoma; fibrosarcoma;
KW renal cell carcinoma; immunotherapy; therapy; retrovirus; vector.
XX
OS Synthetic.
XX
PN WO9624676-A1.
XX
PD 15-AUG-1996.
XX
PF 07-FEB-1996; 96WO-US001787.
XX

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
DB 1 GGGGGGGGGGGGGG 15
|||||
RESULT 6
AAR95067
ID AAR95067 standard; peptide; 15 AA.
XX
AC AAR95067;
XX
DT 18-AUG-1996 (first entry)
XX
DE scFv spacer peptide.
XX
KW Nucleic acid transfer system; gene transfer; gene therapy;
KW cell targeting; multidomain protein; vector; cancer; scFv;
KW single chain antibody.
XX
OS Synthetic.
XX
PN WO9613599-A1.
XX
PD 09-MAY-1996.
XX
PF 31-OCT-1995; 95WO-EP004270.
XX
PR 01-NOV-1994; 94EP-00810627.
XX
PA (WELS/) WELS W.
XX
PI Wels W, Fominaya J;
XX
DR WPI; 1996-239505/24.
XX
PT Nucleic acid transfer system for gene therapy, e.g. against cancer -
PT includes toxin translocation domain to target nucleic acid to specific
PT cell.
XX
PS Example 5; Page 29; 106pp; English.

```

```

PR 08-FEB-1995; 95US-00385335.
XX
XX (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
XX
XX Lieschke GJ, Mulligan RC;
XX
XX WPI; 1996-384448/38.
XX
DR N-PSDB; AAT35195, AAT35196, AAT35202, AAT35203.
XX
XX New DNA encoding fusion protein, esp. contg. IL-12 p35 and p40 subunits -
XX for treatment of established tumours or prevention of tumour
XX establishment.
XX
XX Claim 2; Page 69; 118pp; English.
XX
XX Peptide linkers (Gly4Ser)2Ser, (Gly4Ser)3Ser, (Gly4Ser)3 and (Gly6)Ser
XX (AAR99242-45) are used to join the subunits of novel dimeric or
XX multimeric fusion proteins. They have been utilised in the prodn. of
XX bioactive interleukin-12 (IL-12) fusion proteins, linking mouse/human IL-
XX 12 p35 subunit (see also AAR99246) to mouse/human IL-12 p40 subunit
XX (AAR99247). DNA encoding such constructs can be incorporated into a
XX retroviral vector (see also AAT35198) to allow dimeric IL-12 prodn. in
XX transfected cells. Tumour cells (esp. CMS-5, B16 or renal carcinoma
XX cells) secreting IL-12 dimer can be used to reduce the size of
XX established tumours and/or increase survival time, esp. in cases of
XX melanoma, fibrosarcoma and renal cell carcinoma
XX
XX Sequence 15 AA;
XX
XX Query Match 100.0%; Score 84; DB 2; Length 15;
XX Best Local Similarity 100.0%; Pred. NO. 0.006;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GGGGGGGGGGGGGG 15
XX DB 1 GGGGGGGGGGGGGG 15
XX |||||
XX
XX RESULT 6
XX AAR95067
XX ID AAR95067 standard; peptide; 15 AA.
XX
XX AC AAR95067;
XX
XX DT 18-AUG-1996 (first entry)
XX
XX DE scFv spacer peptide.
XX
XX KW Nucleic acid transfer system; gene transfer; gene therapy;
XX KW cell targeting; multidomain protein; vector; cancer; scFv;
XX KW single chain antibody.
XX
XX OS Synthetic.
XX
XX PN WO9613599-A1.
XX
XX PD 09-MAY-1996.
XX
XX PF 31-OCT-1995; 95WO-EP004270.
XX
XX PR 01-NOV-1994; 94EP-00810627.
XX
XX PA (WELS/) WELS W.
XX
XX PI Wels W, Fominaya J;
XX
XX DR WPI; 1996-239505/24.
XX
XX PT Nucleic acid transfer system for gene therapy, e.g. against cancer -
XX PT includes toxin translocation domain to target nucleic acid to specific
XX PT cell.
XX
XX PS Example 5; Page 29; 106pp; English.

```

XX A spacer peptide (AAR95067) is used to link the light chain variable
 CC domain to the heavy chain variable domain of a single chain recombinant
 CC antibody (scFv). It allows correct folding of an antigen binding domain
 CC present in the variable domains. The scFv is derived from hybridoma FRP5,
 CC which produces monoclonal antibody against the HER2 antigen of human
 CC tumour cells. It forms the ligand domain of a multidomain protein (see
 CC also AAR95053 and AAR95056-58) that is used with an effector nucleic acid
 CC in a novel nucleic acid transfer system suitable for gene therapy. The
 CC ligand domain has a target cell recognition function and allows cellular
 CC internalization of the multidomain protein/nucleic acid complex
 XX

XX Sequence 15 AA;
 SQ

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 1 GGGGGGGGGGGGGG 15

RESULT 7
 AAW09323
 ID AAW09323 standard; peptide; 15 AA.

XX AC AAW09323;
 XX DT 10-JUN-1997 (first entry)
 XX DE Peptide linker arm #1.

XX Chimaeric; bispecific; DNA binding domain; trans; activator; repressor;
 KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
 KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;
 KW hyperproliferation; p53; tumour; oligomerisation.

XX Synthetic.
 XX WO9630512-A1.
 XX 03-OCT-1996.

XX 29-MAR-1996; 96WO-FR000477.
 XX 31-MAR-1995; 95FR-00003841.
 XX (RHON) RHONE-POULENC RORER SA.

XX Bracco L, Schweighoffer F, Tocque B;
 XX WPI; 1996-455359/45.

XX Conditional gene expression system triggered by e.g. infection or hyper-
 PT proliferation - comprises novel bispecific proteins having DNA-binding
 PT domain and second domain specific for trans-activator or repressor, for
 PT gene therapy.

XX Claim 23; Page 45; 81pp; French.

XX The invention relates to novel chimaeric, bispecific proteins which
 CC comprise: (a) a DNA binding domain and (b) a domain which binds a trans-
 CC activator (TA), trans-repressor (TR) or their complexes, which are
 CC characteristic of a physiological or physiopathological state. The novel
 CC chimaeric, bispecific proteins allow expression of a therapeutic protein.
 CC (e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
 CC antibodies) to be regulated in response to particular conditions.
 CC Examples include making the protein responsive to the presence of
 CC particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7
 CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
 CC will be expressed in those cells infected by that pathogen. Similarly,
 CC where the chimaeric protein responds to a cellular protein typical of a

CC hyperproliferative state (esp. wild-type and mutant p53), expression can
 CC be restricted to tumour cells. The sequence presented here is an example
 CC of a peptide linker "arm" which connects the DNA binding domain to the TA
 CC binding domain
 XX

XX Sequence 15 AA;
 SQ

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 1 GGGGGGGGGGGGGG 15

RESULT 8
 AAY49219
 ID AAY49219 standard; peptide; 15 AA.

XX AC AAY49219;
 XX DT 07-FEB-2000 (first entry)
 XX DE Sequence of a linking peptide.

XX Monoclonal antibody; MAb; 1A7; GD2; immune response; melanoma;
 KW neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;
 KW tumor-associated antigen.

XX Synthetic.
 XX US5977316-A.
 XX 02-NOV-1999.

XX 16-JAN-1996; 96US-00591196.
 XX 17-JAN-1995; 95US-00372676.
 XX (KENT) UNIV KENTUCKY.

XX Foon KA, Chatterjee SK, Chatterjee M;
 XX WPI; 1996-354530/35.

XX Monoclonal antibody 1A7 and related polynucleotide(s) and polypeptide(s)
 PT - useful to treat or palliate a GD2-associated disease, e.g. melanoma and
 PT glioma.

XX Disclosure; Col 24; 74pp; English.

XX The invention provides a monoclonal antibody (MAb) designated 1A7, which
 CC elicits an anti-GD2 (tumor-associated antigen) immunological response in
 CC humans. MAb 1A7 has defined light and heavy chain variable region
 CC sequences. The MAb 1A7 and polypeptides can be used for eliciting an anti
 CC -GD2 immune response. The polypeptides can also be used for detecting or
 CC purifying anti-GD2 antibody. The products can be used for treating GD2 -
 CC associated diseases, e.g. melanoma, neuroblastoma, glioma, soft tissue
 CC carcinoma, and small cell carcinoma. They can be used for palliating the
 CC disease or for reducing the risk of recurrence

XX Sequence 15 AA;
 SQ

Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
 |||||
 Db 1 GGGGGGGGGGGGGG 15

RESULT 9
AAW10295
ID AAW10295 standard; peptide; 15 AA.
XX
AC AAW10295;
XX
DT 11-SEP-1997 (first entry)
XX
DE Peptide linker for soluble fused MHC heterodimer:peptide complex.
XX
KW Soluble; fusion; major histocompatibility complex; MHC; heterodimer;
complex; linker; antigen; binding groove; tolerance; autoantigen;
KW disease; insulin dependent; diabetes mellitus; IDDM; antagonist; T cell;
KW energy; presenting cell.
XX
OS Synthetic.
XX
PN W09640944-A2.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US010102.
XX
PR 07-JUN-1995; 95US-00480002.
PR 07-JUN-1995; 95US-00482133.
PR 07-JUN-1995; 95US-00483241.
PR 27-OCT-1995; 95US-0005964P.
XX
PA (ZYMO) ZYMOGENETICS INC.
PA (ANER-) ANERGEN INC.
XX
PI Kindsvogel W, Reich EP, Gross JA, Deshpande S, Sheppard PO;
XX
DR WPI; 1997-052337/05.
XX
PT Novel fused major histocompatibility complex:antigenic peptide complex -
PT useful to induce tolerance to an autoantigen-related disease e.g. insulin
-dependent diabetes mellitus.
XX
PS Claim 7; Page 137; 142pp; English.
XX
CC A novel soluble fused major histocompatibility complex (MHC)
heterodimer:peptide complex, comprises DNA encoding 1st and 2nd MHC
domains, linked by DNA encoding a 5-25 residue linker, e.g. the present
peptide, and a DNA encoding an antigenic peptide able to associate with a
peptide binding groove of the MHC molecule, linked in frame to the DNA
encoding the 2nd domain by a DNA encoding a 5-25 residue linker. The
complex can be used to induce immunological tolerance in adults
susceptible to, or suffering from an autoantigen related disease, e.g.
insulin dependent diabetes mellitus (IDDM), by antagonising the binding
of particular T cells and antigen presenting cells, to induce anergy
(immunological non-responsiveness) in the targeted T cell. As the
heterodimers and corresponding antigen are permanently linked into a
single chain, obviating the requirement for complex heterodimer
truncation or formation, the complex eliminates inefficient and non-
specific peptide loading
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGGGGGGGGGGGS 15
DB 1 GGGGGGGGGGGGGGS 15
RESULT 10
AAW35984
ID AAW35984 standard; peptide; 15 AA.
XX
AC AAW35984;

XX
DT 11-MAR-1998 (first entry)
XX
DE Peptide linker SEQ ID NO:18 from US5683983.
XX
KW Interleukin 5; IL-5; receptor; inflammatory disease; eosinophil; asthma;
beta-adrenergic agonist; corticosteroid; treatment; mimetic;
primary library.
KW
OS Synthetic.
XX
PN US5683983-A.
XX
PD 04-NOV-1997.
XX
PF 07-JUN-1995; 95US-00484083.
XX
PR 07-JUN-1995; 95US-00484083.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Chen M, Sloan D, Barrett RW, England BP, Schatz PJ;
XX
DR WPI; 1997-549007/50.
XX
PT Treatment of disorders mediated by interleukin-5 - by administering
peptide that binds to IL-5 receptor, for treatment of inflammatory
diseases.
XX
PS Disclosure; Col 41-42; 38pp; English.
XX
CC A novel method has been developed for treating a disorder mediated by IL-
5 (interleukin-5). The method comprises administering a peptide that
binds to the IL-5 receptor and comprises the following amino acid
sequence, and dimers and oligomers of this: Cys X1 X2 Trp X3 Arg Cys X4
X5 Cys; where X1 = Gly, Ile, Val or Tyr; X2 = Asp or Glu; X3 = Ala or Val
; X4 = Gln or Pro; and X5 = Ala, Glu, Lys, Met, Asn, Ser or Thr, where
one or more of the CONH linkages may be replaced by a CH2OC(O)NR,
phosphonate, CH2SO2NR, CH2NR, C(O)NR6 or NHCONH linkage, R = H or lower
alkyl and R6 = lower alkyl; the N-terminal group = NR1, NRCOR, NRCOOR,
NRSO2R, NHCONHR, succinimido or NHC(O)CH2Ar; R1 = H or lower alkyl and Ar
= phenyl optionally mono-, di- or tri-substituted by lower alkyl, lower
alkoxy, Cl and Br; the C-terminal group is COR2, R2 = OH, lower alkoxy or
NR3R4, R3 and R4 = H or lower alkyl, or the N atoms of the NR3R4 group
can optionally be part of the amine group of the N-terminus of the
peptide so as to form a cyclic peptide. The present sequence represents a
peptide linker. The peptide causes the production and accumulation of
eosinophils in tissues. It may be used for treating IL-5-mediated
inflammatory disorders, preferably of the respiratory tract, especially
asthma, optionally together with a beta-adrenergic agonist, an
antiinflammatory corticosteroid or ipratropium bromide
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGGGGGGGGGGGS 15
DB 1 GGGGGGGGGGGGGGS 15
RESULT 11
AAW87784
ID AAW87784 standard; peptide; 15 AA.
XX
AC AAW87784;
XX
DT 11-MAY-1999 (first entry)
XX
DE Antibody-beta-lactamase fusion protein spacer peptide #1.
XX

KW Fusion protein; antibody; light chain; heavy chain; variable region;
 KW melanoma-associated antigen; beta-lactamase; cytotoxic agent; prodrug;
 XX tumour cell.

XX Synthetic.

XX WO9850432-A1.

XX 12-NOV-1998.

XX 30-APR-1998; 98WO-US008840.

XX 07-MAY-1997; 97US-0045888P.

XX 30-APR-1998; 98US-00070637.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Siemers NO, Yarnold S, Senter PD;

XX WPI; 1999-070092/06.

XX New fusion polypeptide of antibody variable regions and beta-lactamase -
 PT are targeted to melanoma-associated antigens and used to generate
 PT cytotoxic agents from prodrugs, at tumour cells.

XX Disclosure; Page 6; 50pp; English.

XX The invention relates to a new fusion polypeptide comprising an antibody
 CC light and heavy chain variable regions specific for a melanoma-associated
 CC antigen (Ag) linked to a beta-lactamase (BL). This peptide represents a
 CC spacer peptide used to separate the heavy and light chains of the
 CC antibody. The fusion protein is used to deliver cytotoxic agents to
 CC tumour cells; it binds to a tumour cell Ag and converts an administered
 CC prodrug to the active form

XX Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.006;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGGSGGGSGGGGS 15

Db 1 GGGSGGGSGGGGS 15

RESULT 12

AA43414

ID AAY43414 standard; peptide; 15 AA.

XX AAY43414;

XX 20-DEC-1999 (first entry)

XX Peptide SEQ ID NO:13.

XX Angiogenic homology region; AHR; thrombospondin 1; TSP-1; angiostatin;
 KW endostatin; anticancer; antiangiogenic; cancer; cardiovascular disease;
 KW obesity; osteoarthritis; duodenal ulcer; abnormal neovascularisation;
 KW wound healing; arteriosclerosis; ischaemic limb; ischaemic myocardium;
 KW diabetes mellitus; blood vessel occlusion.

XX Synthetic.

XX WO9848924-A1.

XX 30-SEP-1999.

XX 23-MAR-1999; 99WO-US006334.

XX 24-MAR-1998; 98US-00046737.

XX (CHIL-) CHILDRENS MEDICAL CENT.

PA (YISS) YISSUM RES & DEV CO.

XX Ben-Sasson SA;

XX WPI; 1999-591075/50.

XX New angiogenic peptide derivatives, used for treating e.g. cancer,
 PT cardiovascular diseases, obesity, osteoarthritis, duodenal ulcers,
 PT abnormal neovascularization and for wound healing.

XX Disclosure; Page 59; 62pp; English.

XX The present invention specifically describes peptide derivatives
 CC comprising an angiogenic homology region (AHR) of endostatin. The peptide
 CC derivatives can be used for modulating angiogenesis in humans and
 CC animals. The peptides can be used to treat a wide variety of disease
 CC conditions, including cancer, cardiovascular diseases (e.g.

CC arteriosclerosis, ischaemic limbs and ischaemic myocardium), obesity,
 CC osteoarthritis, duodenal ulcers, abnormal ocular neovascularisation,
 CC associated e.g. with diabetes mellitus, and to promote wound healing or
 CC to stimulate the growth of new blood vessels to bypass, e.g. blood vessel
 CC occlusions. The peptide derivatives can also be used for the production of
 CC of antibodies. The multivalent ligands may enable the administration of
 CC lower doses in order to achieve therapeutic efficacy, as compared with a
 CC univalent peptide chain. In addition, they can have long in vivo
 CC lifetimes and good biodistribution when administered orally or
 CC parenterally. The present sequence represents a peptide used in the
 CC exemplification of the present invention

XX Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.006;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGGSGGGSGGGGS 15

Db 1 GGGSGGGSGGGGS 15

RESULT 13

AA43328

ID AAY33328 standard; protein; 15 AA.

XX AAY33328;

XX 29-NOV-1999 (first entry)

XX E6-sFv peptide linker fragment.

XX Cytotoxic; RNase; ribonuclease; pancreatic; antibody; light chain;
 KW heavy chain; cell surface marker; treatment; tumor; viral infection;
 KW parasite infection; immune dysfunctional cell; autoimmune disease;
 KW contraceptive; cell separation; transplantation; bone marrow ablation;
 KW leukemia cell; T-cell; graft-versus-host disease; ss.

XX Synthetic.

XX US955073-A.

XX 21-SEP-1999.

XX 09-JUL-1997; 97US-00891848.

XX 20-APR-1990; 90US-00510696.

XX 22-OCT-1991; 91US-0079195.

XX 04-FEB-1993; 93US-00014082.

XX 22-SEP-1993; 93US-00125462.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Rybak SM, Newton DL, Nicholls PJ, Youle RJ;

DR WPI; 1999-560488/47.
 XX Recombinantly fused pancreatic RNase-targeting proteins useful for
 PT treating tumors, infections, immune or autoimmune disorders and as a
 PT contraceptive.
 XX Example 3; Col 57-58; 47pp; English.
 PS
 CC This invention describes a novel nucleic acid construct comprising
 CC sequences encoding functional pancreatic RNase and a second protein
 CC (preferably the light and heavy chains of an antibody) which binds a
 CC specific cell surface marker on a target cell and functions as a
 CC cytotoxic agent. The products can be used for selectively killing cells
 CC expressing a specific surface marker. They can be used for treating
 CC tumors or infected cells (e.g. cells infected by viruses (especially
 CC latent or chronic virus infections, such as human immunodeficiency virus
 CC (HIV)-1, Epstein-Barr virus, herpes viruses (herpes simplex types I and
 CC II), hepatitis viruses (B, non-A-non-B, and delta), herpes zoster,
 CC cytomegalovirus)) and cells infected with parasites (such as the malaria
 CC parasite)). They can also be used for treating immune dysfunctional cells
 CC in immune and autoimmune diseases. Additionally, they may be used as
 CC contraceptives. Finally they can also be used for cell separation in
 CC vitro by selectively killing unwanted types of cells (e.g. in bone
 CC marrow) prior to transplantation into a patient undergoing marrow
 CC ablation by radiation or for killing leukemia cells or T-cells that would
 CC cause graft-versus-host disease. This sequence represents a E6-sfv linker
 CC peptide which is used in the method of the invention
 XX
 SQ Sequence 15 AA;
 Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGGGGGGGGGGGG 15
 DB 1 GGGGGGGGGGGGGG 15
 |||||
 RESULT 14
 AAY03763
 ID AAY03763 standard; peptide; 15 AA.
 XX
 AC AAY03763;
 XX
 DT 10-JUN-1999 (first entry)
 XX
 DE Linker peptide in a single-gene encoding an Ig-like molecule.
 XX
 KW Signal-chain; immunoglobulin-like molecule; Ig; monoclonal antibody;
 KW MAB CC49; human; colorectal; carcinoma; biosensor; gene therapy;
 KW linker peptide.
 XX
 OS Synthetic.
 XX
 PN US5892019-A.
 XX
 PD 06-APR-1999.
 XX
 PF 01-SEP-1994; 94US-00299999.
 XX
 PR 15-JUL-1987; 87US-00073685.
 PR 02-JUL-1990; 90US-00547336.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Kashmiri SVS, Shu L, Schlom J;
 XX
 DR WPI; 1999-253934/21.
 XX
 CC Nucleic acid encoding single-chain antibody-like molecule.
 XX
 PS Claim 28; Col 9; 15pp; English.

XX
 CC The invention relates to a single gene encoding a signal-chain
 CC immunoglobulin (Ig)-like molecule that has binding affinity for an
 CC antigen. The Ig-like molecule comprises (a) the binding part (Bp1) of the
 CC light chain variable region (VL); (b) the binding part (BP2) of the heavy
 CC chain variable region (VH); (c) at least one linker connecting Bp1 and
 CC BP2; and (d) a polypeptide comprising a modified heavy chain constant
 CC region; where the Ig-like polypeptide binds to both TAG-72 and LSI747
 CC antigens and specifically binds to an epitope also bound by a monoclonal
 CC antibody CC49 (ATCC CRL9459). The Ig-like polypeptides have the same
 CC specificity, binding properties and cytotoxicity of the parent monoclonal
 CC antibody, so are useful for therapy and diagnosis, specifically of human
 CC (colorectal) carcinoma and their metastases, also in biosensors, for
 CC imaging or for purification. Nucleic acids encoding the Ig-like
 CC polypeptides may be used in gene therapy. Use of the nucleic acid
 CC eliminates (a) the need to deliver two genes and (ii) problems of
 CC inefficient assembly, associated with expression of complete antibodies.
 CC The Ig-like polypeptide induce little, if any, anti-murine antibody
 CC response, and can be used to transfect cells, e.g. tumour-infiltrating
 CC lymphocytes, ex vivo for subsequent delivery to a tumor
 XX
 SQ Sequence 15 AA;
 Query Match 100.0%; Score 84; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGGGGGGGGGGGG 15
 DB 1 GGGGGGGGGGGGGG 15
 |||||
 RESULT 15
 AAY21600
 ID AAY21600 standard; peptide; 15 AA.
 XX
 AC AAY21600;
 XX
 DT 20-MAR-2003 (revised)
 DT 11-AUG-1999 (first entry)
 XX
 DE BP-919566 peptide Seq ID No: 23.
 XX
 KW Antimicrobial; oligopeptide; cecropin P1; microbial pathogen; magainin;
 KW plant pathogen; food additive; preservative; cosmetic; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN EP919566-A2.
 XX
 PD 02-JUN-1999.
 XX
 PF 31-JAN-1992; 98EP-00121780.
 XX
 PR 01-FEB-1991; 91US-00649784.
 PR 31-JAN-1992; 92EP-00101616.
 XX
 PA (ENIE) ENICHEM SPA.
 XX
 PI Mapelli C, Dugas De Robertis C, Stahl GF, Bascomb NF;
 PI Swerdloff MD, Williams JI, Everett NP;
 XX
 DR WPI; 1999-304793/26.
 XX
 PT New oligopeptides containing at least two antimicrobial peptides, useful
 PT for protecting plants against microbial pathogens.
 XX
 PS Disclosure; Fig 1; 67pp; English.
 XX
 CC The invention relates to antimicrobial peptides including reverse
 CC antimicrobial peptides, antimicrobial oligopeptides and other
 CC antimicrobial compositions such as cecropin P1. The antimicrobial
 CC oligopeptides are active against at least one microbial pathogen, and

CC comprise at least one of a first and one of a second peptide monomer,
CC interconnected directly through a peptide bond via the N and C terminals,
CC or indirectly through a disulfide bond or via bridges. At least one of
CC the first and second monomers confers activity. Oligopeptides connected
CC by bridges do not have the structure of Magainin Pre-pro protein. The
CC antimicrobial peptides are used for providing protection to plants
CC against plant pathogens, thus enhancing crop yields. The peptides are
CC also useful for treatment of human or animal disease, as an additive to
CC foods for preservation, or as a preservative in cosmetics and
CC pharmaceuticals. Unlike prior art antimicrobial peptides Magainins 1 and
CC 2, the new antimicrobial peptides don't have undesirable properties, are
CC not subject to extensive proteolytic degradation, are not phytotoxic to
CC the cell, and have a broader range of activity. (Updated on 20-MAR-2003
CC to correct PF field.) (Updated on 20-MAR-2003 to correct PR field.)
XX

SQ Sequence 15 AA;
Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006; 0; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGGGGGG 15
| | | | | | | | | |
Db 1 GGGGGGGGGGGGGG 15

Search completed: April 20, 2004, 10:25:02
Job time : 61.6 secs